

**UNIVERSIDAD AUTONOMA DE MADRID**

**ESCUELA POLITECNICA SUPERIOR**



**Grado en Ingeniería de Tecnologías y Servicios de  
Telecomunicación**

**TRABAJO FIN DE GRADO**

**Modelado de trastornos neurodegenerativos a través de  
sistemas afectivos.**

**Javier Vinuesa Solana**  
**Tutor: Aythami Morales Moreno**  
**Ponente: Julián Fierrez Aguilar**

**Julio 2020**



# **Modelado de trastornos neurodegenerativos a través de sistemas afectivos.**

**AUTOR: Javier Vinuesa Solana**  
**TUTOR: Aythami Morales Moreno**

**Dpto. Tecnología electrónica y de las Comunicaciones**  
**Escuela Politécnica Superior**  
**Universidad Autónoma de Madrid**  
**Julio de 2020**





## Resumen (castellano)

El objetivo de este trabajo de fin de grado es estudiar como el aprendizaje profundo basado en dominios afectivos puede ayudar en diferentes tareas relacionadas con el análisis de expresiones faciales. Una de estas tareas será la detección de la enfermedad neurodegenerativa del Parkinson.

Para conseguir nuestro objetivo empezamos el trabajo recopilando información sobre el estado del arte de los temas más importantes que íbamos a tratar: el análisis facial, dominios afectivos y Enfermedad de Parkinson. La literatura relacionada indica que los adultos mayores con Enfermedad de Parkinson tienen una menor expresividad facial, conocida como hipomimia. Para detectar la hipomimia y ser capaces de clasificar entre pacientes sanos y pacientes con la enfermedad, proponemos una serie de experimentos basados en los modelos de aprendizaje profundo para el análisis de expresiones faciales.

Los experimentos se dividen en dos fases. En primer lugar se utilizarán dos bases de datos afectivas (Affectnet y CFEE) y redes neuronales pre-entrenadas (VGG y Resnet) para reconocimiento facial. Estos modelos se adaptarán al dominio afectivo a través de las bases de datos propuestas y las populares técnicas de *Transfer Learning*. Una vez obtenidos los resultados, se escogerá el modelo que mejor se adapte al escenario de Parkinson. Aprovechando las características aprendidas por el modelo, vamos se aplicará nuevamente la técnica de *Transfer Learning* en este caso para pasar del dominio afectivo al del Parkinson, quedándonos con todas las capas del modelo menos la última y añadiéndole un clasificador de dos salidas. Con este nuevo modelo vamos a realizar la segunda fase, la clasificación de una base de datos con pacientes sanos y pacientes con la Enfermedad de Parkinson. Gracias a este segundo experimento el modelo aprenderá características relacionadas con los pacientes con la Enfermedad de Parkinson.

Finalmente se realizan las conclusiones acerca de lo que el modelo generado va a poder aportar y ayudar a la medicina y se proponen distintos temas para realizar un trabajo futuro acerca de esta investigación.

## Abstract (English)

The objective of this Bachelor Thesis is to study how deep learning based on affective domains can help in different tasks related to the analysis of facial expressions. One of these tasks will be the detection of the neurodegenerative disease of Parkinson's.

To achieve our goal we started the work by gathering information on the state of the art of the most important topics we are going to deal with: facial analysis, affective domains and Parkinson's Disease. The related literature indicates that older adults with Parkinson's disease have reduced facial expression, known as hypomimia. To detect hypomimia and be able to classify between healthy patients and patients with the disease, we propose a series of experiments based on deep learning models for the analysis of facial expressions.

The experiments are divided into two phases. First, two affective databases (Affectnet and CFEE) and pre-trained neural networks (VGG and Resnet) will be used for facial recognition. These models will be adapted to the affective domain

through the proposed databases and the popular Transfer Learning techniques. Once the results are obtained, the model that best suits the Parkinson's scenario will be chosen. Taking advantage of the characteristics learned by the model, we will again apply the Transfer Learning technique in this case to move from the affective domain to that of Parkinson's, keeping all the layers of the model except the last one and adding two-exit classifier. With this new model we are going to carry out the second phase, the classification of a database with healthy patients and patients with Parkinson's disease. Thanks to this second experiment the model will learn characteristics related to patients with Parkinson's disease.

Finally, the conclusions are drawn about what the model generated will be able to contribute and help medicine and different topics are proposed for future work on this research.

## **Palabras clave (castellano)**

Aprendizaje profundo, redes neuronales, redes convolucionales, dominios afectivos, aprendizaje máquina, análisis facial, enfermedades neurodegenerativas, Enfermedad de Parkinson.

## **Keywords (inglés)**

Deep learning, neural networks, convolutional networks, affective domains, machine learning, facial analysis, neurodegenerative disease, Parkinson's disease.





## ***Agradecimientos***

*Primero de todo, quiero dar las gracias a mi tutor, Aythami, por su gran ayuda y compromiso durante este año, pero sobre todo, por despertar en mí un gran interés en las tecnologías utilizadas en este proyecto.*

*Agradecer también a mis padres y mi familia por su gran apoyo y confianza durante estos años, no solo académicamente, sino en todos los aspectos de mi vida. Nunca lo hubiese conseguido sin vosotros.*

*Gracias a mis amigos de toda la vida, tanto areñales como madrileños, por el apoyo y los buenos ratos que me habéis dado. Habéis hecho que este viaje se haga mucho más ameno. Gracias también a mis compañeros de clase, en especial al grupo del Faro, por la ayuda incondicional y todo lo que hemos pasado. Sin vosotros el día a día no hubiese sido lo mismo.*



## INDICE DE CONTENIDOS

1	Introducción.....	1
1.1	Introducción.....	1
1.2	Objetivos.....	2
1.3	Motivación.....	3
1.4	Organización de la memoria.....	3
2	Estado del arte .....	5
2.1	Introducción.....	5
2.2	¿Qué se puede aprender de las caras?.....	5
2.3	Pasos para el análisis facial .....	5
2.4	DCNN's para el análisis facial .....	6
2.5	Reconocimiento de emociones .....	6
2.5.1	Introducción.....	6
2.5.2	Representación de emociones.....	7
2.5.3	Enfoque multimodal .....	7
2.5.4	Detección temprana de Parkinson .....	7
2.5.4.1	FACS en Parkinson .....	7
2.5.4.2	Enmascaramiento facial o Hipomimia .....	9
3	Desarrollo .....	11
3.1	Bases de datos utilizadas .....	11
3.1.1	Introducción.....	11
3.1.2	Compound Facial Expressions of Emotion (CFEE).....	11
3.1.2.1	Emociones Básicas .....	12
3.1.2.2	Emociones Compuestas.....	13
3.1.3	AffectNet .....	14
3.1.3.1	Datos Utilizados .....	15
3.1.4	Base de datos de Parkinson de la Universidad de Antioquia .....	16
3.2	CNN's Utilizadas.....	18
3.2.1	¿Qué es el Transfer Learning?.....	18
3.2.1.1	Fine Tuning .....	19
3.2.2	VGG16 .....	20
3.2.3	ResNet50 .....	21
3.2.4	ResNet entrenada desde 0.....	23
3.3	Experimentos realizados.....	23
3.3.1	CFEE: 8 Emociones .....	23
3.3.2	CFEE: 15 Emociones .....	24
3.3.3	CFEE: 26 Emociones .....	24
3.3.4	Affectnet: 8 Emociones .....	24
3.3.5	Experimentos para la detección de Parkinson .....	25
4	Integración, pruebas y resultados .....	27
4.1	Resultados en reconocimiento de emociones .....	27
4.1.1	Affectnet: 8 Emociones .....	27
4.1.2	CFEE: 8 Emociones .....	28
4.1.3	CFEE: 15 Emociones .....	29
4.1.4	CFEE: 26 Emociones .....	30
4.2	Resultados en detección de Parkinson.....	31
5	Conclusiones y trabajo futuro.....	33
5.1	Conclusiones.....	33

5.2 Trabajo futuro .....	34
6 Referencias .....	35
6.1 Referencias bibliográficas .....	35

## INDICE DE FIGURAS

FIGURA 2-1 EJEMPLO DE ALGUNAS ACTION UNITS. [1] .....	8
FIGURA 3-1. EMOCIONES BÁSICAS. [2] .....	12
FIGURA 3-2. EMOCIONES COMPUESTAS. [3] .....	14
FIGURA 3-3. CATEGORÍAS DE AFFECTNET. [4] .....	15
FIGURA 3-4. EJEMPLO DE TRANSFER LEARNING. [5] .....	18
FIGURA 3-5. EJEMPLO DE FINE TUNNING. [6] .....	19
FIGURA 3-6. CAPAS DEL MODELO VGG16. [7] .....	20
FIGURA 3-7. ESTRUCTURA DEL MODELO VGG16. [8] .....	21
FIGURA 3-8. EJEMPLO DE SKIP CONECTION. [9] .....	21
FIGURA 3-9. ARQUITECTURA DEL MODELO RESNET50. [10] .....	22
FIGURA 3-10. ARQUITECTURA RESNET DEPENDIENDO DEL NÚMERO DE CAPAS. [11] .....	22

## INDICE DE TABLAS

TABLA 3-1. TABLA DE INFORMACIÓN DE LOS DATOS UTILIZADOS POR LAS DOS BASES DE DATOS AFECTIVAS .....	16
TABLA 3-2. TABLA DE INFORMACIÓN SOBRE LOS USUARIOS DE LA BASE DE DATOS DE PARKINSON .....	17
TABLA 4-1. TABLA DE RESULTADOS DE AFFECNET 8 EMOCIONES .....	27
TABLA 4-2. TABLA DE RESULTADOS DE CFEE 8 EMOCIONES .....	28
TABLA 4-3. TABLA DE RESULTADOS DE CFEE 15 EMOCIONES .....	30
TABLA 4-4. TABLA DE RESULTADOS DE CFEE 26 EMOCIONES .....	31
TABLA 4-5. TABLA DE RESULTADOS DE 256 CARACTERÍSTICAS. ....	32

TABLA 4-6. TABLA DE RESULTADOS DE 2048 CARACTERÍSTICAS .....	32
--	----

# 1 Introducción

---

## 1.1 Introducción

La enfermedad de Parkinson (EP) es un trastorno neurológico caracterizado por deficiencias motoras y no motoras que afecta a entre el 1 y el 2 por ciento de las personas mayores de 65 años [1]. Los déficits motores incluyen rigidez, inestabilidad postural, temblores y disartria; y los no motores incluyen depresión, ansiedad, trastornos del sueño y lentitud de pensamiento. Además de la extensa lista de síntomas, la mayoría de los pacientes con EP también exhiben dificultades para expresar emociones o expresiones específicas en sus rostros. Los posibles signos de esas anomalías incluyen un movimiento del músculo facial sin amplitud, una mayor apertura de los ojos, una boca medio abierta y un parpadeo más lento. Todos estos fenómenos en su expresión facial se agrupan en la literatura y se denominan hipomimia [2], que es el resultado de las deficiencias motoras a nivel de los músculos faciales. Normalmente no se nota en las primeras etapas de la EP, pero una vez que se produce un deterioro significativo o un movimiento facial muy reducido, que puede dar lugar a rostros inexpresivos, provocará una capacidad muy limitada para sonreír, para expresar otras emociones o sentimientos como la felicidad, la tristeza, la ira, el miedo, el asco y la sorpresa [3]. El principal efecto de estos impedimentos son las dificultades que tendrán las personas con EP para la comunicación no verbal, lo que puede provocar un aislamiento social a medio y largo plazo.

La evaluación clínica de los pacientes con EP es realizada principalmente por neurólogos expertos según la Sociedad de Trastornos del Movimiento – Escala de Calificación de la Enfermedad de Parkinson Unificada (MDSUPDRS) [4]. Esta escala es el estándar global para la evaluación clínica de los pacientes con EP y considera tanto los síntomas motores como los no motores. La evaluación de la escala MDS-UPDRS varía entre 0 y 4, donde 0 significa completamente sano y 4 significa completamente dañado. Según las directrices de la Sociedad de Trastornos del Movimiento, los 5 niveles con los que se evalúa la hipomimia pueden utilizarse para evaluar las expresiones faciales en los pacientes con EP [4]. En la siguiente lista se indica la correspondencia entre los niveles utilizados y su significado en términos de evaluación de la expresión facial:

- 0) → Normal: Expresión facial normal
- 1) → Ligera: Muy pocas facies enmascaradas, que se manifiestan sólo por la disminución de la frecuencia de parpadeo.
- 2) → Leve: Además de la disminución de la frecuencia de parpadeo, también se presentan facies enmascaradas en la parte inferior del rostro, es decir, menos movimientos alrededor de la boca, como una sonrisa menos espontánea, pero los labios no se separan.
- 3) → Moderado: facies enmascaradas con los labios separados en algún momento cuando la boca está en reposo.

4) → Severo: facies enmascaradas con los labios separados la mayor parte del tiempo cuando la boca está en reposo.

La evaluación neurológica depende en gran medida de la pericia del clínico, lo que provoca una variabilidad y un posible sesgo en el procedimiento de calificación. Por lo tanto, se requiere el desarrollo de sistemas computarizados para apoyar objetivamente la evaluación de la progresión de la enfermedad. Hay varias contribuciones en estado del arte donde se introducen sistemas computarizados para evaluar diferentes aspectos de los pacientes con Parkinson, incluyendo el habla [5], la marcha [6], la escritura [7], el movimiento de las manos [8] y la expresión facial [9]. Entre todos ellos, la expresión facial y la hipomimia parecen ser los aspectos menos cubiertos. El reconocimiento de expresiones faciales (REF) se refiere a la evaluación de la capacidad de los pacientes con EP para reconocer efectivamente diferentes expresiones o emociones cuando miran a la cara. La evaluación de la expresividad facial (FEE) se refiere a la capacidad del paciente de producir diferentes expresiones faciales o emociones. Ambos aspectos tienen un papel muy importante en la interacción social y la comunicación no verbal. El primero ha sido estudiado durante varios años, principalmente por psicólogos en diferentes trabajos y los principales descubrimientos se resumen en el estudio recientemente publicado por Argaudetal [10]. Por otra parte, aunque la FEE se ha convertido en un campo popular entre los ingenieros e informáticos, ha abierto espacio a la investigación en diferentes aplicaciones de la informática afectiva. La base principal es crear modelos computacionales con la capacidad de reconocer, interpretar y procesar las emociones humanas, lo que hace que la interacción entre el hombre y la computadora sea cada vez más eficiente. El análisis de los sentimientos y la computación afectiva han sido continuamente estudiados desde el siglo XX, ayudando en el desarrollo de los sistemas de visión artificial [11], [12], en la creación de entretenimiento [13], y en el desarrollo de sistemas para ayudar a diferentes áreas de la medicina incluyendo la neurología [14], [15], [16]. El trabajo se centra en el estudio de la FEE en pacientes con EP. [17]

## **1.2 Objetivos**

El objetivo principal del trabajo es estudiar como el aprendizaje profundo basado en dominios afectivos puede ayudar en diferentes tareas relacionadas con el análisis de los gestos faciales. Una de las tareas en las que puede ayudar es en el ámbito de la neurología, para la detección y evaluación de la Enfermedad de Parkinson.

Para esto, el objetivo principal se va a dividir en dos sub-objetivos. El primero va a ser el análisis y entrenamiento de modelos de reconocimiento de emociones mediante redes neuronales convolucionales y dos bases de datos afectivas. El segundo consiste en realizar una clasificación entre pacientes sanos y pacientes con la Enfermedad de Parkinson a partir del modelo que mejor se ajuste a las características de los enfermos de Parkinson.

Una vez se tienen conclusiones sobre los dos sub-objetivos, se van a poner en común para sacar una resolución acerca del objetivo principal y con esto aportar lo que se pueda al estudio de dominios afectivos mediante aprendizaje profundo y a la detección de Parkinson.

### **1.3 Motivación**

El aprendizaje profundo o *Deep Learning* es una de las técnicas del aprendizaje automático más utilizadas y que mejores resultados está obteniendo en infinidad de tareas. Una de las tareas más estudiadas y trabajadas es la conocida como análisis facial.

Dentro del análisis facial se encuentran los dominios afectivos. Estos constituyen todo lo relacionado con el análisis de emociones. El análisis de las emociones se puede utilizar en infinidad de tareas, entre ellas las aportaciones médicas, como por ejemplo para el estudio de enfermedades neurodegenerativas, dentro de las que se encuentra la Enfermedad de Parkinson.

Para el estudio y la detección temprana del Parkinson se han realizado muchas aportaciones, sobre todo relacionadas con el habla o la escritura, sin embargo, no hay muchas relacionadas con el dominio afectivo de pacientes con la enfermedad. Recientemente se han hecho investigaciones que concluyen que un estudio del dominio afectivo de los pacientes puede ayudar a detectar la hipomimia y con esto detectar la enfermedad de Parkinson.

La motivación de este trabajo se basa en aportar un pequeño granito de arena a la detección y evaluación de esta enfermedad mediante un largo estudio de las redes neuronales profundas y los dominios afectivos.

### **1.4 Organización de la memoria**

La memoria consta de los siguientes capítulos:

- Capítulo 1: Introducción, Objetivos y Motivación del proyecto.
- Capítulo 2: Estado del Arte. Aquí hablaremos sobre las aportaciones e investigaciones recientes en los temas de reconocimiento facial, sistemas afectivos y Enfermedad de Parkinson.
- Capítulo 3: Desarrollo. Aquí se describen los experimentos que realizamos en el proyecto. También se hablará de las bases de datos y redes neuronales pre-entrenadas que hemos utilizado.
- Capítulo 4: Integración, pruebas y resultados de los experimentos realizados en el proyecto.
- Capítulo 5: Conclusiones y trabajo futuro.





## 2 Estado del arte

---

### 2.1 Introducción

Los recientes avances en las redes neuronales convolucionales profundas (DCNN's) han mostrado impresionantes mejoras de rendimiento en varios problemas de detección/reconocimiento de objetos. Esto ha sido posible gracias a grandes bases de datos etiquetadas y a una mejor comprensión de la cartografía no lineal entre las imágenes y las etiquetas de clase, así como a la facilidad de acceso a potentes unidades de procesamiento gráfico (GPU's). Estos avances en el aprendizaje profundo también han mejorado la capacidad de las máquinas para comprender los rostros y ejecutar automáticamente las tareas de detección de caras, estimación de pose, localización de puntos de referencia y reconocimiento de caras a partir de imágenes y videos sin restricciones.

### 2.2 ¿Qué se puede aprender de las caras?

El análisis facial es un desafío típico en visión por computadora del cual se lleva investigando activamente más de dos décadas [18]. El objetivo es extraer del rostro la mayor cantidad de información posible, como la ubicación, la pose, el género, la identificación, la edad la emoción, etc. Las aplicaciones de esta tecnología incluyen la detección e identificación de una persona de interés a partir de videos de vigilancia, la autenticación activa de los *smartphones*, las transacciones de pago mediante la biometría facial, análisis médico, los coches inteligentes, etc. Además, ha habido un interés creciente en el reconocimiento y la verificación de rostros a partir de imágenes y videos sin restricciones que también implican sub-tareas como la detección de caras, localización de puntos de referencia faciales, etc.

### 2.3 Pasos para el análisis facial

Normalmente se necesitan tres módulos para realizar un análisis facial en una imagen. Primero, se aplica un detector de rostros para localizar las caras en imágenes o videos. Para un sistema robusto, el detector de rostros debe ser capaz de detectar caras con diferente pose, iluminación y escala. Además, la ubicación y el tamaño de las cajas delimitadoras de las caras deben determinarse con precisión para que contengan una cantidad mínima de fondo. En segundo lugar, se utiliza un detector de puntos de referencia importantes como los centros de los ojos, la punta de la nariz, las esquinas de la boca, etc. Estos puntos se utilizan para alinear las caras a las coordenadas canónicas normalizadas para mitigar los efectos de la rotación y el escalado. En tercer lugar se extrae de la cara alineada un descriptor de rasgos que codifica la información facial.

## 2.4 DCNN's para el análisis facial

Para abordar estos desafíos los investigadores han empleado el aprendizaje profundo para obtener las características requeridas para el análisis facial. Las redes neuronales convolucionales profundas (DCNN's) han demostrado ser efectivas para las tareas de análisis de imágenes. En los últimos años, las DCNN's se han utilizado para resolver numerosos problemas de visión por computadora como el reconocimiento de objetos [19], [20] y la detección de objetos [21], [22]. Una DCNN típica está formada por una jerarquía de capas convolucionales con una función de activación de unidad lineal rectificadora, que consiste en millones de parámetros entrenables. Debido a las buenas representaciones obtenidas, las DCNN han sido aplicadas recientemente para la detección de rostros [23], [24], [25], la localización de puntos de referencia [23], [25], [26], y la extracción de características, es decir para los tres módulos del análisis facial.

Un factor clave que contribuye a este excepcional rendimiento es la disponibilidad de un gran número de bases de datos de caras etiquetadas, tales como, CASIA-WebFace [27], MegaFace [28], [29], MS-Celeb-1M [30], y VGGFace [31]. La gran cantidad de datos de entrenamiento representa variaciones significativas en la pose, iluminación, expresión y oclusión, lo que permite a las DCNN's ser robustas a estas variaciones y extraer características significativas requeridas para la tarea.

## 2.5 Reconocimiento de emociones

### 2.5.1 Introducción

El análisis de la expresión del rostro es un tema que ha sido ampliamente estudiado por los investigadores relacionados con el análisis facial. La mayoría de los trabajos existentes se han centrado en el análisis de la expresión facial para predecir las emociones [32], [33]. La base de estos métodos es el Sistema de Codificación de Acción Facial (FACS) [34], que codifica la expresión facial utilizando un conjunto de movimientos específicos localizados en la cara, llamados *Action Units*. En la **Figura 2-1** podemos ver algún ejemplo de *Action Units*. Estos enfoques basados en la cara [32], [33] suelen utilizar características basadas en la geometría facial o en la apariencia para describir el rostro. Después las características extraídas se utilizan para reconocer las *Action Units* y las emociones básicas propuestas por Ekman y Friesen [35]: ira, asco, miedo, felicidad, tristeza y sorpresa. En la actualidad los sistemas más novedosos en cuanto al reconocimiento de emociones por el análisis de expresión facial, utilizan las redes neuronales convolucionales profundas (DCNN's) para reconocer las emociones o las *Action Units*.

El Reconocimiento de Emociones en el Contexto de la Escena y al Análisis del Sentimiento de la imagen son problemas diferentes que comparten algunas características. El Reconocimiento de Emociones tiene como objetivo identificar las emociones de las personas en las imágenes. El Análisis del Sentimiento de la imagen consiste en predecir lo que una persona sentirá al observar una imagen. Esta imagen no tiene que contener necesariamente una persona. Cuando una imagen contiene una persona, puede haber una diferencia entre las emociones experimentadas por la persona en la imagen y las emociones sentidas por los observadores de la imagen. Por ejemplo si vemos la imagen de un atleta que ha

perdido un partido, podemos reconocer que el atleta se siente triste. Sin embargo, un observador de la imagen puede sentirse feliz si el observador es fan del equipo que ganó el partido.

## 2.5.2 Representación de emociones

En cuanto a la representación de emociones, algunos de los trabajos recientes basados en la expresión facial [36] utilizan dimensiones continuas del Modelo de Estado Emocional VAD. El modelo VAD describe las emociones utilizando 3 dimensiones numéricas: Valencia (V), que mide cuán positiva o agradable es una emoción, que va de negativa a positiva; Arousal (A), que mide el nivel de agitación o excitación de la persona, que va de inactiva/en calma a agitada/lista para actuar; y Dominio (D) que mide el nivel de control que una persona siente de la situación, que va de sumisa/sin control a dominante/en control. Por otro lado, Du et al [37] propuso un conjunto de 21 categorías de emociones faciales, definidas como diferentes combinaciones de las emociones básicas, como ‘felizmente disgustado’ o ‘felizmente sorprendido’. Con esta categorización los autores pueden dar un gran detalle sobre la emoción expresada.

## 2.5.3 Enfoque multimodal

Aunque la investigación en el reconocimiento de las emociones desde una perspectiva de visión informática se centra principalmente en el análisis de la cara, hay algunos trabajos que también consideran otras pistas visuales adicionales o enfoques multimodales. Por ejemplo, en [38] la ubicación de los hombros se utiliza como información adicional para el reconocimiento de las emociones básicas en el rostro del nutriólogo. En general, Schindleretal [39] utilizó la postura del cuerpo para reconocer 6 emociones básicas, realizando experimentos sobre una pequeña base de datos de gente con Alzheimer. Mou et al [40] presentaron un sistema de reconocimiento de emociones, a partir de imágenes fijas de grupos de personas, reconociendo la excitación y la valencia a nivel de grupo a partir de la combinación de información facial, corporal y contextual.

















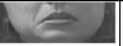






## 2.5.4 Detección temprana de Parkinson

La enfermedad de Parkinson (EP en adelante) es una de las enfermedades neurodegenerativas más comunes, que afecta aproximadamente al 1% de los individuos mayores de 60 años y al 2-4% de los mayores de 75 años, causando una discapacidad progresiva que exhibe la mayor pérdida de años (en cuanto a calidad de vida) de entre las 29 principales enfermedades crónicas [41] [42].

### 2.5.4.1 FACS en Parkinson

El Sistema de Codificación de Acción Facial (FACS) es una herramienta psicológica social que se utiliza para medir la presencia, la intensidad y las características temporales de 44 movimientos musculares, llamados *Action Units* (AU). Algunos de estos los podemos ver en la **Figura 2-1**. El uso anterior del FACS en personas con EP ha investigado principalmente las diferencias en las expresiones faciales espontáneas y deliberadas y ha investigado el uso de la sonrisa de Duchenne entre las personas con EP [43]; [44]. La sonrisa de Duchenne implica

la activación del músculo elevador de la mejilla (AU 6) que crea la apariencia de patas de gallo alrededor de los ojos además de la sonrisa de la boca (AU 12) [45] y a menudo se percibe que muestra más felicidad que sonrisas sin la activación del elevador de la mejilla [46]. Utilizando el sistema FACS para medir la sonrisa, se descubrió que una muestra de cuatro personas con EP producía menos sonrisas espontáneas de Duchenne que los sujetos sanos [43], sin embargo, en una muestra de seis personas con EP que no producían sonrisas espontáneas de Duchenne pudieron producir una sonrisa de Duchenne cuando se les pidió que lo hicieran deliberadamente [44]. La sonrisa de Duchenne es posiblemente el medio más potente para comunicar una emoción positiva, y la pérdida de la capacidad de producir sonrisas de Duchenne podría ser bastante estigmatizante.

Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
					
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink
Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
					
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
					
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
					
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

**Figura 2-1** Ejemplo de algunas Action Units. **Fuente:** <https://www.pinterest.es/pin/821273682017790454/>

#### **2.5.4.2 Enmascaramiento facial o Hipomimia**

##### **2.5.4.2.1 ¿Qué es?**

La expresión de las emociones es importante para mantener relaciones sociales que eviten la soledad y fomenten el apoyo social a lo largo de la vida. Los adultos mayores que padecen la enfermedad de Parkinson (EP) suelen tener una menor expresividad facial, lo que se denomina enmascaramiento facial o hipomimia, debido a una menor capacidad para activar la acción motora en la cara [47] [48]. La hipomimia (o enmascaramiento facial) es un signo de la EP que a menudo se presenta en sus primeras etapas [49]. El síndrome se caracteriza por una marcada disminución de los gestos expresivos de la cara, incluyendo los movimientos de la frente que acompañan al habla y las expresiones faciales emocionales [50]. Los movimientos de cejas –contracciones muy breves (~50 ms) de los músculos de la parte superior de la cara que ocurren durante el habla y que parecen añadir énfasis semántico, a menudo están ausentes. Además, la hipomimia se manifiesta comúnmente sólo en un lado de la cara [51].

Se sabe poco acerca de cómo se relaciona el enmascaramiento facial con la experiencia de estigmatización y bienestar social u otros síntomas motores y no motores de la EP. Para comenzar a comprender estas preguntas, se necesita una comprensión completa de como la EP afecta a todas la áreas del rostro de la persona con la enfermedad. Wu [52], comenzó su trabajo creando mapeos que incluían la presencia y los aspectos temporales de 11 *Action Units* codificadas a partir de videos de 30 segundos de personas con EP que veían videos repugnantes y neutrales, pero éstos eran de repuestas emocionales inducidas más que de expresiones faciales espontáneas y no incluían la intensidad del movimiento o la relación entre las activaciones de las unidades de acción a través de la cara.

La cartografía sistemática de la expresión facial a partir de una narración espontanea aumenta la validez ecológica y permite generalizar la medición a la vida social real de la persona.

##### **2.5.4.2.2 Evaluación**

Las personas, incluidos los médicos, suelen formar impresiones del rostro como expresiones emocionales Gestalt, sin reconocer explícitamente las relaciones subyacentes entre las diferentes regiones del rostro. Esto se refleja en la actual medida clínica de enmascaramiento facial que utiliza expresiones generales para clasificar a las persona con EP. Estas medidas son: enmascaramiento facial leve, mínimo, moderado o severo [53]. Estas impresiones gestálticas son heurísticos útiles para hacer impresiones de la cara [54], pero a menudo es la activación de los movimientos musculares en conjunción con los movimientos en otras regiones de la cara lo que produce sutiles distinciones entre las percepciones de las expresiones emocionales. Por ejemplo, una sonrisa en la parte inferior del rostro se considera más genuina cuando va acompañada de mejillas levantadas en la mitad superior del rostro [46], y una boca abierta con las cejas levantadas se percibe a menudo como una sorpresa, mientras que una boca abierta con la ceja fruncida se percibe más a menudo como ira [55]. Dado que las diferentes regiones de la cara son independientes desde el punto de vista motor y neural, para evaluar como el enmascaramiento facial influye en los movimientos coordinados de la cara, es

necesario medir la activación de los músculos faciales con la especificidad de cada movimiento individual [51].

#### **2.5.4.2.3 Detección de Hipomimia para diagnóstico de Parkinson**

El efecto de la hipomimia es un síntoma motor secundario de la enfermedad de Parkinson. La hipomimia es un grado reducido de expresión facial que puede ser causado por una deficiencia motora (por ejemplo, debilidad o parálisis de los músculos faciales). Como se informó en [56], los experimentos revelaron que los participantes en la EP mostraron una menor expresividad facial espontánea en las situaciones experimentales y tuvieron más dificultades que los controles para plantar expresiones emocionales e imitar movimientos faciales no emocionales.

Por lo tanto, la detección de la hipomimia puede ser utilizada para el diagnóstico de Parkinson. En [57] se desarrollan métodos para detectar y evaluar automáticamente la gravedad de la hipomimia mediante el seguimiento de las expresiones faciales que hacen los pacientes, mientras ven los videoclips. Específicamente, las herramientas de *Machine Learning* se emplean usando mediciones de un sensor de profundidad que permite un seguimiento bastante preciso de los movimientos faciales. Se obtienen resultados alentadores que demuestran el concepto de que la evaluación automática de la hipomimia puede ser lo suficiente fiable como para ser útil para la detección clínica precoz de la hipomimia relacionada con la EP.

Según [58], se ha prestado poca atención a este aspecto clínico relevante de la enfermedad y todavía faltan métodos de evaluación y enfoques terapéuticos validados. En este trabajo intentaremos estimar los síntomas de la hipomimia a partir de tres tipos de bases de datos: dos afectivas y otra de pacientes con EP. [59], [60], [61], [62]

## 3 Desarrollo

---

### 3.1 Bases de datos utilizadas

#### 3.1.1 Introducción

Las bases de datos o *datasets* son una parte fundamental del análisis de imágenes con redes neuronales. Mucha gente cree que tener una buena base de datos consiste en tener una base de datos con una cantidad enorme de imágenes pero esta afirmación es solo verdad en parte, ya que no siempre es así. Para tener buenos resultados con tu red neuronal, es necesario tener cierto número de imágenes para poder extraer todas las características disponibles. Es cierto, que con 10 o 15 imágenes por categoría, los resultados que obtendremos no serán demasiado buenos pero dependiendo del problema, si este no es muy complejo, con cientos o unos pocos de miles de imágenes se pueden llegar a obtener grandísimos resultados. No es necesario tener bases de datos con millones de imágenes para obtener buenos resultados.

Otra parte fundamental en las bases de datos es la división de estas, en grupos de entrenamiento, test y validación. Esto es debido a que nunca puedes testear un modelo de redes neuronales con los mismos datos que has usado para entrenarla. Si utilizas para evaluar tu modelo las imágenes a partir de las cuales se han obtenido los pesos de sus capas (imágenes de entrenamiento), obviamente va a obtener buenos resultados. Para evaluar la efectividad real de tu modelo es necesario que este sea evaluado con imágenes que no haya visto nunca (imágenes de test)

En este apartado vamos a ver las bases de datos, a partir de las cuales se han realizado los entrenamientos y evaluaciones de todos nuestros experimentos.

Para nuestro estudio se han utilizado dos bases de datos afectivas distintas. Una con menos imágenes llamada Compound Facial Expressions of Emotion (CFEE) y otra con un gran número de estas llamado AffectNet. También hemos utilizado una base de datos de pacientes sanos y con EP para los experimentos de detección de Parkinson.

#### 3.1.2 Compound Facial Expressions of Emotion (CFEE)

Esta base de datos fue desarrollada por Shichuan Du, Yong Tao, and Aleix M. Martinez, pertenecientes al departamento de ingeniería eléctrica y de computadores de la Ohio State University. Además también pertenecen al Center for Cognitive and Brain Sciences de la misma universidad. Tiene como fecha de publicación febrero de 2014.

Años atrás las bases de datos de emociones solían estar compuestas de las 8 emociones básicas: neutral, contento, triste, enfadado, sorprendido, desprecio, asco y miedo. Sin embargo la particularidad de esta base de datos es que contiene 26 tipos de emociones distintas. Estas emociones se corresponden a tales como ‘sorprendido-feliz’ o ‘sorprendido-miedoso’.



Este tipo de categorización es consistente con la categorización básica. Esto quiere decir que las métricas que nos aparecen a la hora de clasificar emociones como ‘sorprendido’ o ‘feliz’ aparecerán de manera conjunta en la emoción ‘sorprendido-feliz’.

Al ser un tipo de base de datos con emociones compuestas, las cuales no se han explorado mucho, es primordial intentar reducir los efectos debidos a la luz, la pose y otros que no tengan que ver como tal con la expresión. Debido a esto, al contrario que una base de datos *in to the wild*, todas las fotos están tomadas desde el mismo ángulo, con la misma cámara, misma iluminación, etc.

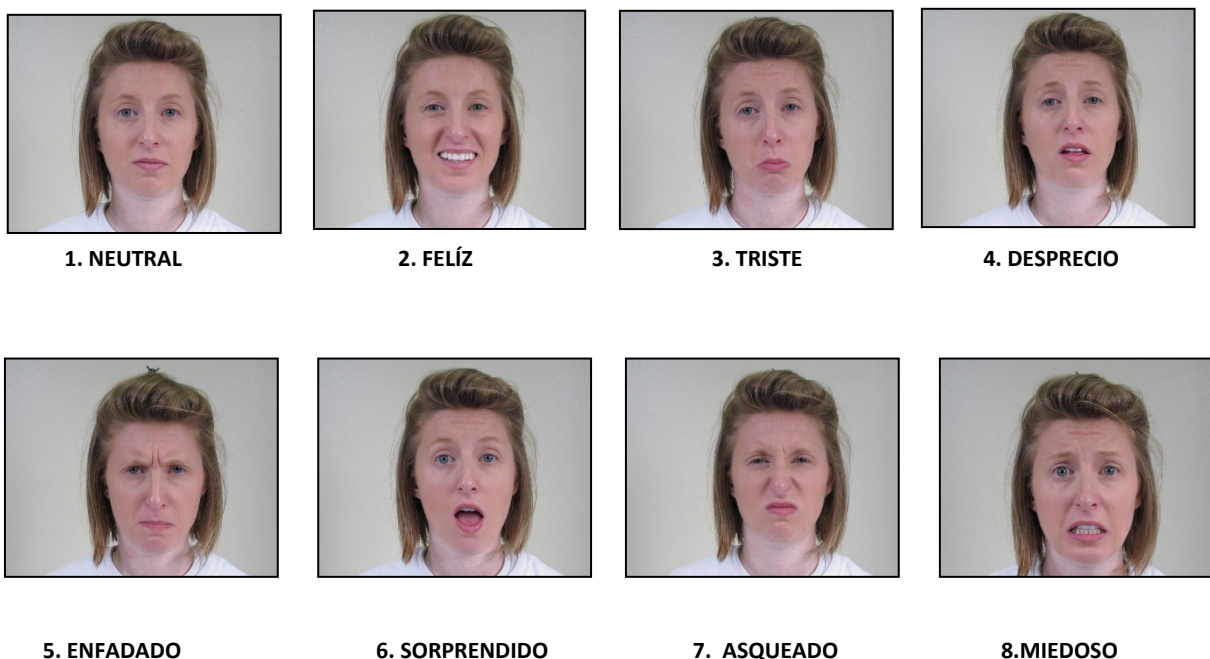
Los sujetos utilizados para realizar la base de datos fueron 230 alumnos (130 mujeres y 100 hombres) de la universidad de Ohio. Entre estos alumnos había representantes de todas las etnias: caucásicos, afro-americanos, asiáticos...

Para categorizar las imágenes se utilizan las mencionadas anteriormente *Action Units*. Esta métricas se utilizan para ‘representar’ los movimientos o contracciones de los músculos de la cara, de tal manera que podamos hacer una clasificación de las imágenes según la expresión que estén representando. Por ejemplo, la *Action Unit* 4, representa la contracción de dos músculos en la parte de debajo de la ceja. Estos se puede ver en la **Figura 2-1**

Por lo tanto esta base de datos se puede dividir en 2 grandes grupos: Emociones Básicas y Emociones Compuestas. [37]

### 3.1.2.1 Emociones Básicas

Se llaman de esta manera porque no se pueden descomponer en etiquetas semánticas más pequeñas. Este grupo está formado por 1840 imágenes, pertenecientes a 8 emociones (230 identidades, 7 emociones por cada uno más la neutral). Estas 8 emociones se corresponden con las que aparecen en la **Figura 3-1**:

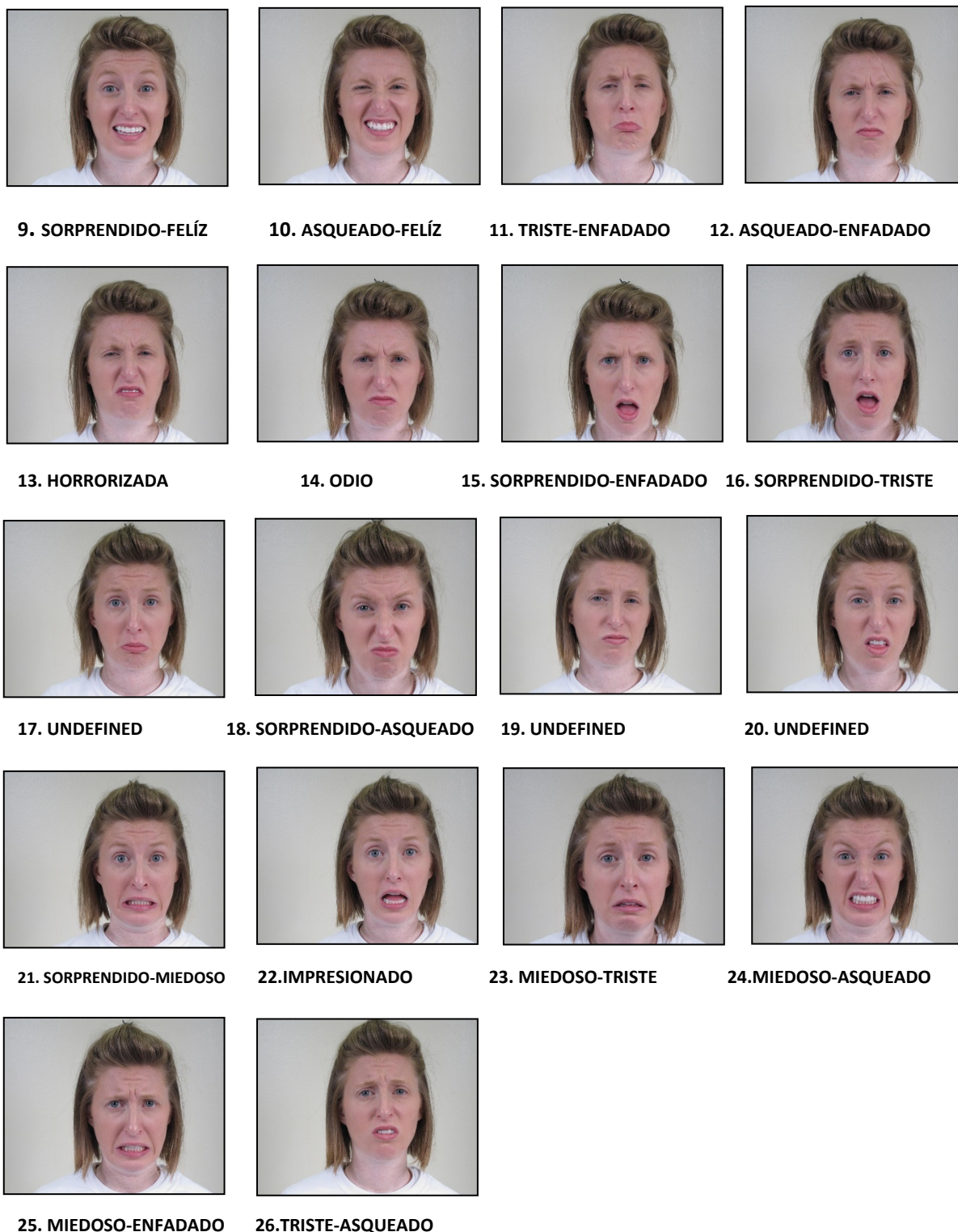


**Figura 3-1.** 8 Emociones Básicas de la base de datos CFEE. **Fuente:** [http://cbcs1.ece.ohio-state.edu/dbform\\_compound.html](http://cbcs1.ece.ohio-state.edu/dbform_compound.html)

### 3.1.2.2 Emociones Compuestas

El segundo grupo de esta base de datos es el formado por las emociones compuestas. Estas emociones se llaman de esta manera porque están formadas por la suma de dos emociones básicas o por combinaciones entre varias de ellas.

Este segundo grupo está formado por 5980 imágenes, correspondientes a 26 categorías (230 identidades con una foto por cada emoción). En estas categorías se incluyen también las emociones básicas, de las cuales no pondremos ejemplo gráfico al ser las mismas que en el grupo anterior. Por lo tanto las emociones compuestas están formadas por las que aparecen en la **Figura 3-2**:



**Figura 3-2.** 18 Emociones compuestas de la base de datos CFEE. **Fuente:** [http://cbcsl.ece.ohio-state.edu/dbform\\_compound.html](http://cbcsl.ece.ohio-state.edu/dbform_compound.html)

### 3.1.3 AffectNet

Affectnet es una base de datos de expresiones faciales desarrollada por Ali Mollahosseini, Behzad Hasani y Mohammad H. Mahoor, todos ellos pertenecientes al IEEE.

La mayoría de bases de datos etiquetadas de imágenes de expresiones faciales *in to the wild* son muy pequeñas. Con *in to the wild* nos referimos a imágenes que han sido tomadas desde diferentes ángulos, diferente iluminación, con distinta cámara..., es decir, todo lo contrario a la base de datos Compound Facial Expressions of Emotion. Con más de un millón de imágenes recolectadas de internet, Affectnet es la mayor base de datos de expresiones faciales *in to the wild*. Esta base de datos fue creada buscando palabras clave en los tres buscadores más famosos (Google, Bing y Yahoo) y posteriormente anotando por expertos en el etiquetado.

Para realizar la búsqueda de estas imágenes, ciertas palabras clave relacionadas con las emociones fueron combinadas con palabras relacionadas con el género, la edad o la etnia obteniendo 362 *strings* en inglés tales como *joyful girl* o *furious young lady*. Estos *strings* fueron traducidos a otros cinco idiomas: español, portugués, alemán, árabe y persa. No todos los *strings* tenían una traducción literal en todos los idiomas por lo que finalmente se consiguieron 1250 consultas de búsqueda para introducir en los buscadores.

Una vez realizadas estas consultas, se obtuvieron alrededor de 1800000 URL's de imágenes. Posteriormente se utilizó el OpenCv Face Recognition para obtener cajas delimitadoras alrededor de las caras. Después de esto se utilizó un algoritmo de alineación facial con el que se extrajeron 66 puntos de referencia. Estos puntos se utilizaron para realizar la clasificación de las imágenes, la cual se hizo de dos maneras: clasificación dinámica y clasificación por categorías.

La clasificación dinámica se basa en el Modelo de Estado Emocional VAD. Este modelo consiste en clasificar las imágenes según su valencia y su excitación (también conocido como *Arousal*). Con la valencia nos referimos a lo positivo o negativo que es un evento y con excitación a lo excitante o tranquilo que es. El valor de la valencia y la excitación estará siempre entre -1 (negativo/tranquilo) y 1 (positivo/excitado). Por ejemplo las imágenes de gente con la emoción 'feliz', tendrán siempre una valencia entre 0 y 1 y una excitación entre -0.2 y 0.5. Sin embargo nosotros no hemos utilizado esta clasificación para nuestros modelos.

La clasificación que hemos utilizado para este proyecto ha sido la clasificación por categorías. Para esta clasificación se definen 11 categorías de imágenes. Estas categorías se componen de las 8 emociones básicas (en realidad 7 más la neutral) y de las categorías: 'None', en la cual se encuentran imágenes que no corresponden a ninguna de las emociones básicas como por ejemplo avergonzado o confundido; 'Non Face' en la cual hay imágenes correspondientes a dibujos animados o que no contienen caras en la imagen; y por último la categoría 'Uncertain' en la que se encuentran imágenes para las que ha habido duda a la hora de clasificar. A las

imágenes de las categorías ‘Non Face’ y ‘Uncertain’ no se les han asignado ni valencia ni excitación. [63]

Los ejemplos de las imágenes según la categoría son los que aparecen en la Figura 3-3



**Figura 3-3.** Las 10 Categorías de la base de datos AffectNet. **Fuente:** <http://mohammadmahoor.com/affectnet/>

Para realizar la clasificación por categorías de todas las imágenes se utilizaron 2 procedimientos. En primer lugar se realizó una clasificación manual por expertos, los cuales iban mirando imagen a imagen y asignándoles una categoría según su criterio propio. Mediante esta clasificación se clasificaron en torno a 420 mil imágenes. La segunda forma de clasificar estas imágenes fue mediante la red neuronal ResNext, la cual clasificaba automáticamente estas imágenes. Esta red neuronal fue entrenada con una base de datos de imágenes clasificadas manualmente, obteniendo un 65 % de precisión. Según este procedimiento se clasificaron alrededor de 550 mil imágenes.

### **3.1.3.1 Datos Utilizados**

Para realizar este proyecto hemos decidido utilizar el grupo imágenes anotadas manualmente ya que tienen una mayor probabilidad de estar etiquetadas correctamente.

De las 11 categorías de imágenes, hemos decidido utilizar solamente las 8 primeras debido a que las imágenes de las categorías ‘None’, ‘Uncertain’ y ‘Non-Face’ no nos aportaban información relativa a nuestro objetivo.

La realidad es que estas 8 clases están bastante desbalanceadas en cuanto al número de imágenes que contienen. Por ejemplo, la categoría ‘Felicidad’ tiene alrededor de 135000 imágenes, mientras que la categoría ‘Desprecio’ tiene solamente 3750 imágenes. Este desbalance es algo lógico al ser imágenes tomadas de internet. Con esto me refiero a que las imágenes que la gente sube a internet suelen estar más relacionadas con emociones como la felicidad o el asombro que con la tristeza o desprecio.

El desbalanceo de clases puede originar varios problemas. Entre estos se encuentran algunos como, ‘La existencia de subclases poco representadas’, que hace que las imágenes de estas clases puedan ser confundidas con datos atípicos o ruido, o también ‘La falta de densidad en los datos de entrenamiento’ que provoca que no se pueda llevar a cabo una generalización de estos datos al no disponer de una densidad suficiente de ellos.

Por este motivo hemos decidido realizar un ajuste para balancear el número de imágenes de cada clase. Este ajuste es conocido como sub-muestreo de la clase mayoritaria. Este método consiste en reducir el número de imágenes por clase al número de imágenes que tenga la clase minoritaria. De esta forma la base de datos utilizada quedaría como 8 categorías con 3750 imágenes cada una. Estas imágenes estarían repartidas en 2250 para entrenamiento, 750 para evaluación y 750 para test. Un resumen de los datos utilizados por las dos bases de datos afectivas se puede ver en la **Tabla 3-1**.

	Nº de Participantes	Nº de Emociones	¿Condiciones controladas?	Nº de Imágenes Total	Nº de Imágenes de Train	Nº de Imágenes de Test	Nº de Imágenes de Val
<b>CFEE</b>	230	26	Sí	5980	3588	1196	1196
<b>AffecNet</b>	Desconocido	8	No	30000	18000	6000	6000

**Tabla 3-1.** Tabla de información de los datos utilizados por las dos bases de datos afectivas

### 3.1.4 Base de datos de Parkinson de la Universidad de Antioquia

Esta base de datos fue creada por el grupo de investigación GITA. Actualmente tiene grabaciones de 24 pacientes sanos y 31 pacientes con Enfermedad de Parkinson. Los videos están grabados a 30 cuadros por segundo y en ambientes controlados, ya que las condiciones de luz y el fondo detrás de la escena no cambian entre los participantes. Además, los pacientes con EP han sido evaluados por un neurólogo experto de acuerdo con la escala MDS-UPDRS. Un resumen de la información de los usuarios tratados se encuentra en la **Tabla 3-1**.



	Pacientes con EP		Pacientes sanos	
	Hombres	Mujeres	Hombres	Mujeres
Nº de Participantes	19	12	12	12
Edad media	70.2 ± 10.4	67.4 ± 10.9	65.3 ± 8.7	65.2 ± 10.1
Rango de Edad. (Máximo y Mínimo)	[52 – 90]	[53 – 87]	[49 – 83]	[49 – 80]
Hace cuantos años fueron diagnosticados	8.7 ± 5.4	15.6 ± 17.3	-	-
Rango de hace cuantos años fueron diagnosticados. (Máximo y Mínimo)	[2 – 20]	[1 – 45]	-	-
MDS-UPDRS	35.4 ± 13.9	29.7 ± 12.3	-	-
Rango MDS-UPDRS. (Máximo y Mínimo)	[16 – 65]	[15 – 54]	-	-

**Tabla 3-2.** Tabla de información sobre los usuarios de la base de datos de Parkinson

Es importante remarcar, como se ve en la **Tabla 3-1**, que la edad media y el rango de edades entre pacientes sanos y pacientes con EP tiene que ser muy similar. Se necesita que los pacientes sanos tengan algo de deterioro debido a la edad. Esto es importante a la hora de evaluar a los pacientes ya que no es lo mismo comparar a una persona sana con 16 años y un paciente de EP con 65 que evaluar a un individuo sano y otro con EP en rangos de edad similares.

La base de datos consta de 8 grabaciones de vídeo con tareas como guiño del ojo derecho, guiño del ojo izquierdo, sonreír, expresar la ira, expresar la sorpresa, una serie de movimientos de cabeza y la lectura de un monólogo de 32 palabras. Para los experimentos de este trabajo utilizamos las grabaciones de expresión de la sonrisa. **En esta memoria no se incluyen imágenes de los participantes en la bbdd por temas de privacidad.**

## 3.2 CNN's Utilizadas

### 3.2.1 ¿Qué es el Transfer Learning?

Una técnica muy común y efectiva para realizar Deep-learning en bases de datos con un número reducido de imágenes es la conocida como *Transfer-Learning* o uso de una red pre-entrenada. Una red pre-entrenada es una red guardada que ha sido previamente entrenada, por lo general, en un problema de clasificación de imágenes con una base de datos gigante. Si la base de datos con la que ha sido previamente entrenada es lo suficientemente grande y generalista, el modelo de red puede actuar de forma generalista para un gran número de problemas de visión por computador, aunque estos problemas sean con clases totalmente distintas a las clases con las que fue previamente entrenado.

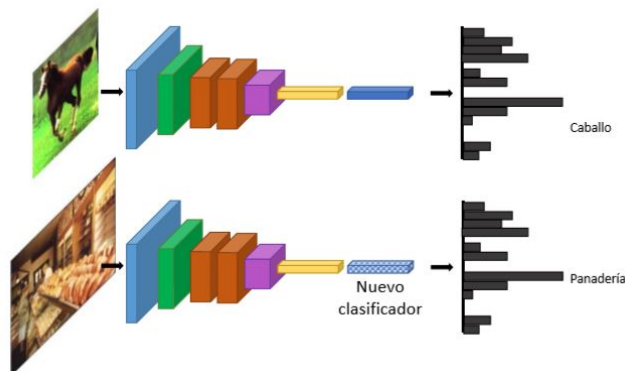
Muchas de las redes han sido previamente entrenadas con la base de datos ImageNet, la cual contiene alrededor de 14 millones de imágenes etiquetadas en 1000 clases diferentes.

Las redes pre-entrenadas más comunes son las familias de redes: VGG, ResNet, Inception, Inception-ResNet y Xception. De esta forma vamos a poder utilizar estas redes para realizar nuestros problemas de clasificación, obteniendo con ellas unos grandes resultados.

Las redes neuronales convolucionales están divididas en dos partes. La primera parte corresponde a las capas Pooling y convolucionales las cuales están conectadas con la segunda parte, correspondiente a las capas densas y el clasificador. La primera parte es conocida como *convolutional-base*.

Cuando realizas problemas utilizando redes pre-entrenadas, utilizas la *convolutional-base* de la red pre-entrenada y lo conectas con un clasificador acorde a tu problema. Al entrenar tu modelo, tendrás que ‘congelar’ los pesos de la *convolutional-base*, de tal manera que se mantengan las características aprendidas con la anterior base de datos. El clasificador, inicializado aleatoriamente, sí que irá aprendiendo las características más específicas de nuestro problema, obteniendo por lo general, unos grandes resultados. [64]

En la **Figura 3-4** se puede ver como una red que ha sido entrenada para clasificar animales, puede ser utilizada para clasificar establecimientos, cambiando únicamente la parte del clasificador.



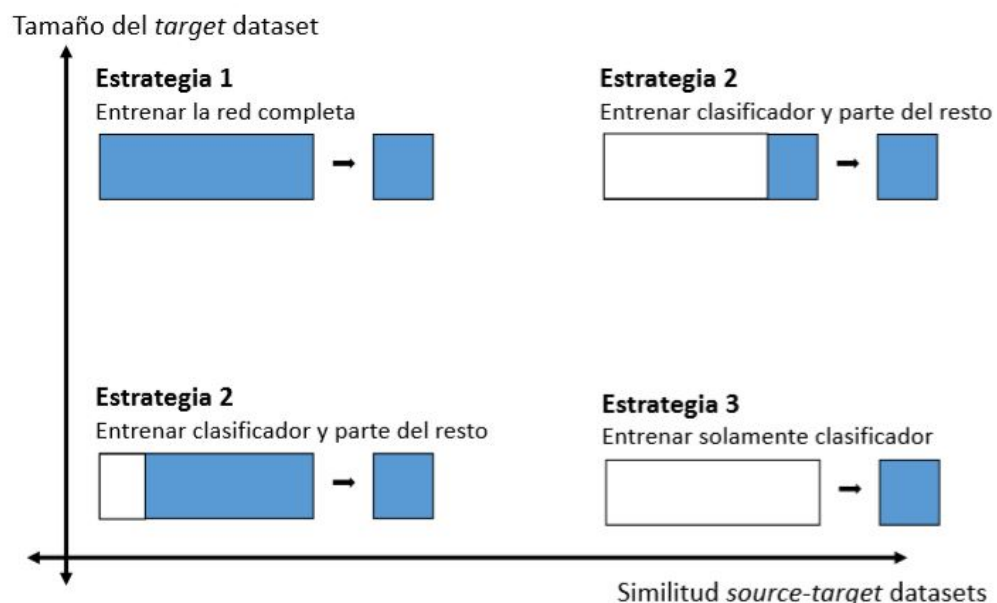
**Figura 3-4.** Ejemplo de dos modelos realizando la técnica de Transfer Learning. **Fuente:** <http://www.cs.cornell.edu/courses/cs4670/2019sp/>

### 3.2.1.1 Fine Tunning

Dentro del Transfer Learning existe otra técnica con la que se pueden mejorar un poco más los resultados. Esta es conocida como *Fine-Tuning* y es la que vamos a utilizar en la mayoría de nuestros experimentos. Esta técnica se basa en realizar primero un entrenamiento como el mencionado anteriormente (clasificador acorde a tu problema y los pesos de la *convolutional-base* congelados) y posteriormente realizar un refinamiento de la red descongelando algunas capas de la *convolutional-base* y volviéndolo a entrenar. De esta manera las capas descongeladas de la *convolutional-base* aprenderán también características relacionadas con nuestro problema. Esto hace que se mejore un pequeño porcentaje la precisión de nuestro modelo.

La cuestión está en cuantas capas de la *convolutional-base* hay que descongelar para obtener los mejores resultados posibles. Eso va a depender de dos factores: la similitud entre las dos bases de datos y del tamaño de la base de datos de nuestro problema (*target dataset*).

Como se puede ver en la **Figura 3-5** si las dos bases de datos se parecen mucho y la base de datos que tenemos para nuestro problema es pequeña, deberemos entrenar solamente el 0. Si se parecen mucho pero la base de datos que tenemos para nuestro problema es grande, realizaremos *Fine-Tuning* descongelando unas pocas capas de la *convolutional-base*. Si por el contrario las bases de datos no se parecen mucho y la base de datos que tenemos es pequeña, debemos descongelar un gran número de capas. Por último si las bases de datos no se parecen en nada y nuestra base de datos es grande, entrenaremos la red desde 0.



**Figura 3-5.** Ejemplo de los tipos de estrategia en Fine Tunning. **Fuente:** San Miguel, J. Carlos (2019). Tema 5: Aprendizaje profundo (Deep Learning) para procesamiento de imágenes Transfer Learning. [Diapositivas de Power Point].

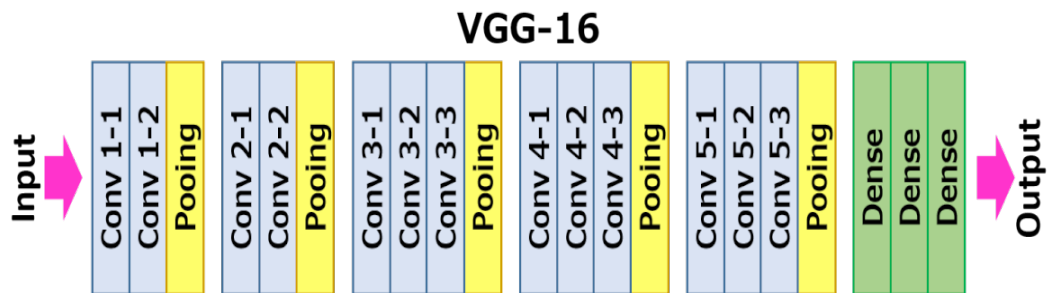


### 3.2.2 VGG16

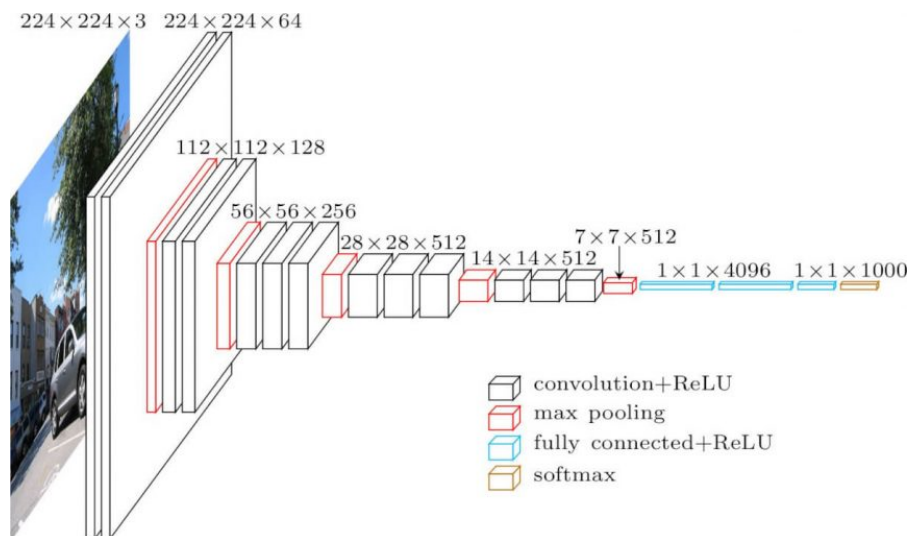
Una de las redes pre-entrenadas que vamos a utilizar es la conocida como VGG16. VGG16 es una red neuronal convolucional desarrollada por K. Simonyan y A. Zisserman pertenecientes al grupo Visual Geometry Group de la Universidad de Oxford. Este modelo fue utilizado para la clasificación de imágenes en diferentes clases. Consiguió alcanzar un 92,7 % de precisión sobre la base de datos ‘ImageNet’, la cual está formada por 14 millones de imágenes pertenecientes a 1000 clases diferentes.

Este modelo es una mejora de la red convolucional AlexNet. Estas mejoras se basan en reemplazar los kernels 11x11 y 5x5 por múltiples 3x3, lo que la convierte en un modelo más profundo, capaz de extraer características más complejas de las imágenes. Esta red fue entrenada durante semanas mediante la NVIDIA Titan Black GPU's. [65]

Vgg16 está formada por 16 capas con sus respectivos pesos. 13 de estas capas son de convolución, 2 son totalmente conectadas y la última es el clasificador, como se puede observar en las figuras (Figura 3-6 y Figura 3-7):



**Figura 3-6.** Estructura por capas del modelo VGG16. **Fuente:** <https://appliedmachinelearning.blog/2019/05/02/building-end-to-end-memory-network-for-question-answering-system-on-babi-facebook-data-set-python-keras-part-2/>

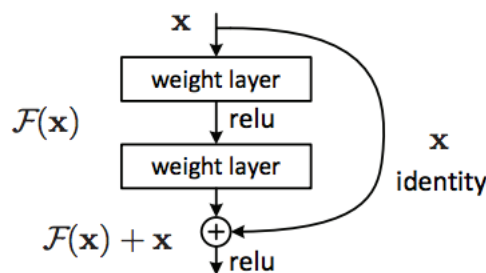


**Figura 3-7.** Estructura del modelo VGG16. **Fuente:** <https://neurohive.io/en/popular-networks/vgg16/>

### 3.2.3 ResNet50

La segunda red con la que vamos a realizar los experimentos y el *Fine-Tuning* es la conocida como ResNet50. ResNet ganó la competición ILSVRC-2015, quedando por delante de la versión VGG de ese año, obteniendo un 2,65% menos de error. Esta competición se basaba en clasificar de manera correcta la base de datos mencionada anteriormente, ‘ImageNet’.

ResNet aparece como una mejora de las redes VGG y a diferencia de estas, es una red recursiva. La diferencia de ResNet con una red convolucional estándar es que tiene unas capas de adición en las que combina la información de capas anteriores, de aquí, lo de recursiva. La recursividad va a provocar que todas las capas obtengan una mayor especialización y por lo tanto obtengamos mejores resultados en tareas más complejas. Esto es una especie de refinamiento. En las redes neuronales convolucionales clásicas la salida de una capa sirve de entrada en la capa siguiente. En las redes recursivas la salida de la capa I, puede ser la entrada de la capa IV. Estas conexiones se conocen como *skip connections* y se pueden observar en la **Figura 3-8**. De esta manera este tipo de redes va a tener muchas más conexiones que una red convolucional clásica. [66], [67]



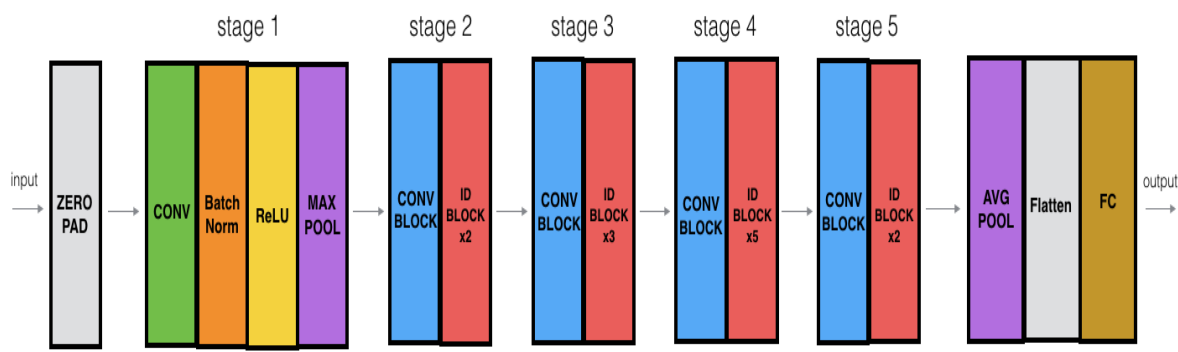
**Figura 3-8.** Ejemplo de skip connection. **Fuente:** <https://kharshit.github.io/blog/2018/09/07/skip-connections-and-residual-blocks>

En vez de utilizar los pesos obtenidos al clasificar la base de datos ‘ImageNet’ hemos utilizado los pesos de la red entrenada con la base de datos ‘VggFace2’. Esta base de datos es una evolución de la base de datos ‘VggFace’. Contiene más de 3,3 millones de imágenes de rostros, pertenecientes a unas 9000 identidades. Esto va a provocar que el modelo ResNet50 ‘aprenda’ características más centradas en caras y no en toda clase de objetos, como ocurre en el modelo de Vgg16 utilizado. [68]

Al igual que en ‘ImageNet’, la base de datos ‘VggFace2’ obtiene mejores resultados de clasificación con la red ResNet que con las redes VGG.

En nuestro caso hemos utilizado el modelo ResNet50, indicando el 50 el número de capas que pueden tener pesos, es decir, parámetros capaces de ser aprendidos.

Como se puede observar en la **Figura 3-9** , estas 50 capas están divididas en 5 *stages*. El primero de estos, es el bloque convolucional 1, el cual contiene solamente una de estas capas. El segundo *stage* es el bloque convolucional 2, el cual está formado por 9 capas con parámetros que pueden ser aprendidos. El bloque convolucional 3 tiene 12, el bloque 4 tiene 18 y por último, el bloque convolucional 5 tiene 9 capas con pesos. A todas estas hay que añadirles la *Fully Connected Layer*, la cual nos da el total de 50.



**Figura 3-9.** Arquitectura por capas del modelo Resnet50. **Fuente:** <https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33>

En la **Figura 3-10** podemos ver la arquitectura ResNet dependiendo del número de capas.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

**Figura 3-10.** Arquitectura ResNet dependiendo del número de capas. **Fuente:** <https://datascience.stackexchange.com/questions/33022/how-to-interpert-resnet50-layer-types/47489>

### 3.2.4 ResNet entrenada desde 0

Hasta ahora todas las redes neuronales convolucionales que hemos mencionado han utilizado las técnicas de *Transfer Learning* y *Fine-Tuning*. Sin embargo, la red presentada en este apartado es una red entrenada desde 0, es decir, las capas no han ‘aprendido’ ninguna característica de imágenes previas, estando todos los pesos de estas inicializados aleatoriamente.

Para la creación de esta red nos hemos basado en el tutorial de [69], en el cual se crea una red recursiva ResNet desde 0. Para ello se crean 3 funciones: una para la generación de bloques residuales (son los típicos de las redes recursivas), otra para realizar las normalizaciones de Relu y Batch y por último la función que define la estructura de nuestra red.

La estructura de esta va a ser: una capa de entrada de imágenes con tamaño (32x32x3), una capa convolucional2D con 64 filtros, (2, 5, 5, 2) bloques residuales con (54, 128, 256 y 512) filtros, una capa AveragePooling2D con pool size = 4, una capa Flatten y una capa densa con (8, 15 o 26 salidas) dependiendo del problema que vayamos a realizar.

## 3.3 Experimentos realizados

Una vez mencionadas las bases de datos y las redes neuronales utilizadas, vamos a hablar de los experimentos realizados.

La primera parte del trabajo consistió en realizar unos experimentos para la clasificación de emociones. Estos se basaban en la utilización de los diferentes modelos (VGG16, Resnet50 y la ResNet entrenada desde 0) para clasificar las imágenes de las bases de datos (AffectNet y CFEE) según la emoción que expresaban los diferentes individuos de las imágenes. Gracias a este experimento los modelos aprenderán características faciales que posteriormente se utilizarán en la tarea de detección de Parkinson.

La segunda parte se basaba en los experimentos relacionados con la detección de la Enfermedad de Parkinson. Estos experimentos van a utilizar el modelo que mejor resultado nos haya dado en la clasificación de emociones, y lo van a utilizar para la detección de Parkinson, mediante la técnica de *Fine-Tuning*.

### 3.3.1 CFEE: 8 Emociones

Este experimento consiste en realizar una clasificación de las primeras 8 emociones de la base de datos Compound Facial Expressions of Emotion mediante las tres redes neuronales. Estas primeras 8 emociones corresponden con las 7 emociones básicas más la neutral.

En este experimento hemos realizado 13 pruebas. Con el modelo ResNet entrenado desde 0 hemos realizado 4 pruebas, en las cuales hemos ido modificando el tamaño de Batch, el número de épocas y la Data-Augmentation. Con VGG16 hemos realizado también 4 pruebas, modificando también el Batch y las épocas. Finalmente con Resnet50 se han realizado 5 pruebas en las que se han ido

modificando el número de capas congeladas en el *fine-tunning* y el tamaño del Batch.

### 3.3.2 CFEE: 15 Emociones

Este experimento es una ampliación del anterior y consiste en realizar una clasificación de las 15 primeras emociones de la base de datos Compound Facial Expressions of Emotion mediante las tres redes neuronales. Para la realización de este, hemos tenido que modificar la capa *softmax* de salida de en los tres modelos, cambiando el clasificador de 8 salidas a 15.

Basándonos en los resultados del primer experimento, hemos realizado 11 pruebas para este. De estas 11 pruebas, 3 corresponden al modelo ResNet entrenado desde 0, 4 al modelo VGG16 y otras 4 al modelo ResNet50. Las pruebas realizadas para este experimento siguen la línea de las realizadas en el experimento de las 8 emociones, con modificaciones de Batch, épocas y capas congeladas, de tal manera que se puedan confirmar los resultados obtenidos con los diferentes modelos y parámetros.

### 3.3.3 CFEE: 26 Emociones

Este experimento es también una ampliación del anterior. En este se va a realizar la clasificación de todas las emociones de la base de datos Compound Facial Expressions of Emotion, es decir, de las 26 emociones. Al igual que en el anterior, para realizarlo hemos modificado el clasificador de los tres modelos, cambiando la capa de salida de 15 salidas a 26.

Para este experimento se van a realizar 11 pruebas, en línea de los otros experimentos, que nos van a servir para confirmar los resultados obtenidos en los otros dos experimentos.

### 3.3.4 Affectnet: 8 Emociones

Este experimento consiste en realizar una clasificación de las 8 primeras emociones de la base de datos Affectnet. La diferencia con los otros experimentos es que estos se realizarán sobre una base de datos *in to the wild*, lo que nos permitirá realizar una comparación entre las dificultades de clasificación de dos bases de datos tan distintas.

Para este experimento se van a realizar 13 pruebas en la misma línea que el experimento de 8 emociones con la base de datos CFEE. La gran diferencia va a ser en la Data Argumentation utilizada. Al ser AffectNet una base de datos *in to the wild* su Data Argumentation incluirá rangos de brillo y rotación, cosa que en CFEE no incluíamos, al ser todas las fotos tomadas en un estudio y con condiciones similares.

### 3.3.5 Experimentos para la detección de Parkinson

Para la realización de estos experimentos se partió del modelo que mejor resultado obtuvo en la clasificación de 26 emociones con CFEE. Utilizamos los resultados con la base de datos CFEE porque sus imágenes se asemejan más a las de la base de datos de Parkinson, siendo estas en un entorno controlado, con iluminación uniforme y pose frontal.

Vamos a realizar la técnica mencionada anteriormente como **Transfer Learning [3.2.1]**. Aprovechando las características faciales aprendidas en la clasificación de emociones, del modelo escogido vamos a coger y congelar todas las capas menos la última, es decir, hasta la capa de 256 características. Posteriormente se le va a añadir un clasificador con dos salidas para realizar la clasificación entre pacientes sanos y pacientes con Enfermedad de Parkinson.

A continuación vamos a realizar el mismo experimento pero esta vez quedándonos con todas las capas del modelo menos las dos últimas, es decir, hasta la capa de 2048 características. Después compararemos en que experimento hemos obtenido mejores resultados.

Como hemos mencionado anteriormente, las personas con la EP tienen una mayor dificultad para realizar expresiones faciales y controlar su musculatura. Por lo tanto, para este experimento vamos a utilizar como imágenes de entrenamiento secuencias de pacientes pasando de un estado neutral a uno sonriendo.

La primera secuencia consta de tres imágenes, neutral, transición-onset, y ápex; la segunda de ápex, transición-offset y neutral; y la tercera es la suma de las dos primeras secuencias. Consideramos ápex a la imagen de la persona sonriendo, transición-onset al paso de neutral a ápex y transición-offset al paso de ápex a neutral.

Al ser un base de datos con pocos participantes, cuando entrenamos los modelos con la base de datos de Parkinson vamos a utilizar la técnica *leave one out*. Esta técnica consiste en entrenar el modelo con todos los datos disponibles menos con uno y evaluas con el usuario que no entrenas. Repites este proceso pero cada vez dejando un dato diferente fuera. Esto consigue que se mejoren los resultados pese a tener un conjunto de datos pequeños.

Gracias a este experimento, el nuevo modelo generado, va a ser capaz de aprender características que le permitan distinguir, con una baja probabilidad de error, entre un paciente sano y un paciente con EP.



## 4 Integración, pruebas y resultados

### 4.1 Resultados en reconocimiento de emociones

#### 4.1.1 Affectnet: 8 Emociones

MODELO	FINE-TUNNING	DATA-ARGUMENTATION	Nº ÉPOCAS	BATCH	TRAIN ACCURACY	TEST ACCURACY
Resnet desde 0	No	Sí	100	20	0.1325	0.116
Resnet desde 0	No	Sí	150	20	0.1290	0.1234
Resnet desde 0	No	Sí	20	128	0.1873	0.1741
Resnet desde 0	No	Sí	150	128	0.5335	0.4394
Resnet desde 0	No	Sí	150	64	0.4433	0.4260
VGG16	Si (78/126 congeladas)	Sí	100 y 200	32	0.4169	0.4776
VGG16	Si (78/126 congeladas)	Sí	200 y 400	32	0.4784	0.4537
VGG16	Si (78/126 congeladas)	Sí	100 y 200	128	0.4621	0.4753
Resnet50	Si (96/126 congeladas)	No	100 y 200	20	0.9750	0.5210
Resnet50	Si (78/126 congeladas)	Sí	100 y 200	20	0.9105	0.5380
Resnet50	Si (52/126 congeladas)	Sí	100 y 200	20	0.9175	0.5540
Resnet50	Si (96/126 congeladas)	Sí	100 y 200	20	0.9065	0.5550
Resnet50	Si (96/126 congeladas)	Sí	100 y 200	32	0.8531	0.5060

**Tabla 4-1.** Tabla de resultados de Affecnet 8 Emociones

La **Tabla 4-1** es la tabla de resultados de los experimentos realizados con los tres modelos sobre la base de datos AffectNet. Dependiendo del modelo y de las experiencias previas, hemos ido modificando ciertos parámetros de los modelos para ir mejorando los resultados de test.

En cuanto al modelo ResNet entrenado desde 0, al no realizarle *Fine Tunning*, nos hemos centrado en el tamaño de Batch y el número de épocas de entrenamiento. Como resultado hemos obtenido que para este modelo, a mayor Batch, mejores resultados se obtenían. Sin embargo si lo entrenabas pocas épocas, no alcanzaba su mejor resultado, como se puede ver en el experimento que consta de 20 épocas. Para alcanzar los mejores resultados con este modelo, se necesitaba tener un Batch



alto (Como 128 o 64) y entrenarlo al menos 70 épocas, que es a partir de la época en la cual los resultados empezaban a ser estables.

En el modelo orientado a la clasificación de objetos, VGG16, sí que se ha realizado *Fine Tunning*. En este modelo hemos modificado las épocas y el Batch, obteniendo resultados muy parejos en todos ellos.

Los mejores resultados se han obtenido con el modelo Resnet50. Esto es lógico ya que es un modelo que ha sido entrenado con una base de datos de caras, al igual que lo es AffectNet. En este caso los parámetros modificados han sido las capas que congelábamos con el *Fine Tunning* y el tamaño de Batch. Todos los resultados son bastante parejos. La parte negativa de este modelo es que tiene un mayor *overfitting* que los otros dos. Sin embargo los resultados de Test son los mejores con diferencia.

#### 4.1.2 CFEE: 8 Emociones

MODELO	FINE-TUNNING	DATA-ARGUMENTATION	Nº ÉPOCAS	BATCH	TRAIN ACCURACY	TEST ACCURACY
Resnet desde 0	No	Sí	100	20	0.1420	0.1442
Resnet desde 0	No	Sí	300	20	0.1465	0.1769
Resnet desde 0	No	Sí	100	64	0.9351	0.5907
Resnet desde 0	No	Sí	150	128	0.9914	0.6479
VGG16	Si (78/126 congeladas)	No	100 Y 200	20	1.000	0.5981
VGG16	Si (78/126 congeladas)	Sí	100 Y 200	20	0.9889	0.6861
VGG16	Si (78/126 congeladas)	Sí	100 Y 200	64	0.8540	0.7138
VGG16	Si (78/126 congeladas)	Sí	25 y 50	128	0.7570	0.6774
Resnet50	Si (78/126 congeladas)	No	100 y 200	20	0.5629	0.1452
Resnet50	Si (78/126 congeladas)	Sí	100 y 200	20	0.9985	0.8017
Resnet50	Si (52/126 congeladas)	Sí	100 y 200	20	0.9980	0.7900
Resnet50	Si (65/126 congeladas)	Sí	100 y 200	20	1.000	0.7730
Resnet50	Si (78/126 congeladas)	Sí	100 y 200	64	0.9998	0.7542

**Tabla 4-2.** Tabla de resultados de CFEE 8 Emociones

La **Tabla 4-2** es la tabla de resultados de los experimentos realizados sobre las 8 primeras clases de la base de datos Compound Facial Expressions of Emotion.

En cuanto al modelo ResNet entrenado desde 0, ocurre como en el experimento previo, a mayor tamaño de Batch, mejores resultados obtiene. Esto se puede observar cuando utilizamos el Batch = 20. Aunque le apliquemos muchas épocas de entrenamiento, nunca alcanza los resultados obtenidos con un Batch de 128.

En el modelo VGG16, el tamaño de Batch no influye tanto. Sí que es verdad que con el Batch = 64 se consiguen mejores resultados que con el Batch = 20, pero esta mejora es mínima. De hecho si aumentas el Batch a 128, los resultados empiezan a empeorar.

Los mejores resultados, al igual que en el experimento anterior, se consiguen con el modelo ResNet50. En los diferentes experimentos que hemos realizado con este modelo, hemos ido modificando el número de capas que aprendían nuevos parámetros, consiguiendo resultados similares con todos ellos. Con este modelo y esta base de datos, se llega a alcanzar una tasa de acierto del 80%. También hemos probado a modificar el Batch, sin embargo, en este modelo no afecta. Los resultados son iguales con Batch = 20 que con Batch = 64.

Los resultados son mejores que en la clasificación de 8 emociones de la base de datos AffectNet. Esto es lógico ya que la base de datos AffectNet no tiene una organización tan completa como CFEE. Con esto me refiero a que la base de datos AffectNet se obtuvo a partir de *strings* introducidos en los buscadores de internet, mientras que la base de datos CFEE se obtuvo con fotos de sujetos específicos a los que se les pedía que posaran para expresar una emoción. Por otra parte, la base de datos CFEE ha sido realizada en un entorno controlado, al contrario que AffectNet. Debido a esto los modelos tendrán mayor facilidad de aprender las características sobre fotos de estudio (CFEE) que sobre imágenes *in to the wild* (AffectNet). La mejora de los resultados es sustancial, pasando del 55,5% de precisión obtenido con AffectNet al 80% obtenido en CFEE.

#### 4.1.3 CFEE: 15 Emociones

MODELO	FINE-TUNING	DATA-ARGUMENTATION	Nº ÉPOCAS	BATCH	TRAIN ACCURACY	TEST ACCURACY
Resnet desde 0	No	Sí	100	128	0.9223	0.3977
Resnet desde 0	No	Sí	100	64	0.7648	0.4025
Resnet desde 0	No	Sí	100	20	0.0600	0.0629
VGG16	Si (78/126 congeladas)	Sí	100 y 200	20	0.5503	0.4244
VGG16	Si (78/126 congeladas)	Sí	100 y 200	32	0.6007	0.4347
VGG16	Si (78/126 congeladas)	Sí	100 y 200	64	0.6479	0.4356
VGG16	Si (78/126	Sí	100 y 200	128	0.6971	0.4468

	congeladas)					
Resnet50	Si (78/126 congeladas)	Sí	100 y 200	20	0.9865	0.5025
Resnet50	Si (96/126 congeladas)	Sí	100 y 200	20	0.9849	0.5360
Resnet50	Si (52/126 congeladas)	Sí	100 y 200	20	0.9894	0.5147
Resnet50	Si (111/126 congeladas)	Sí	100 y 200	20	0.9817	0.4721

**Tabla 4-3.** Tabla de resultados de CFEE 15 Emociones

La **Tabla 4-3** es la tabla de resultados de los experimentos realizados sobre las 15 primeras clases de la base de datos Compound Facial Expressions of Emotion.

Los resultados siguen la línea del experimento anterior. Con el modelo Resnet entrenado desde 0, si el Batch es muy pequeño, los resultados son muy malos. A partir del Batch = 64, los resultados mantienen constantes. Al aumentar el Batch, también aumenta el *overfitting*.

Como se puede observar en la tabla de resultados, el tamaño de Batch no afecta en el modelo VGG16. Aquí los resultados se mantienen bastante constantes en todos los experimentos realizados. Los resultados son mejores que con el modelo ResNet entrenados desde 0.

Como en los anteriores experimentos, los mejores resultados son con el modelo entrenado con imágenes de rostros, Resnet50. Los diferentes experimentos se basaban en modificar el *Fine Tunning* realizado. En todos se consiguieron resultados bastante constantes, siendo el modelo con 96 capas congeladas el mejor de ellos.

Los resultados, como es lógico, son más bajos que en el experimento de clasificación de 8 clases. El mejor resultado que se obtiene nos da un 53,63 % de probabilidad de acierto. Esto es un resultados bastante bueno, teniendo en cuenta que la probabilidad de acierto original era de 6,66% (1/15)

#### 4.1.4 CFEE: 26 Emociones

MODELO	FINE-TUNNING	DATA-ARGUMENTATION	Nº ÉPOCAS	BATCH	TRAIN ACCURACY	TEST ACCURACY
Resnet desde 0	No	Sí	100	128	0.8158	0.2775
Resnet desde 0	No	Sí	100	64	0.5265	0.2561
Resnet desde 0	No	Sí	170	64	0.6228	0.2561
VGG16	Si (78/126	No	200 y 200	20	0.9980	0.303

	congeladas)					
VGG16	Si (78/126 congeladas)	No	200 y 400	20	0.9990	0.304
VGG16	Si (78/126 congeladas)	Sí	200 y 200	20	0.7285	0.342
VGG16	Si (78/126 congeladas)	Sí	100 y 200	32	0.6970	0.340
Resnet50	Si (78/126 congeladas)	Sí	100 y 200	20	0.9679	0.3650
Resnet50	Si (96/126 congeladas)	Sí	100 y 200	20	0.9680	0.3700
Resnet50	Si (111/126 congeladas)	Sí	100 y 200	20	0.96690	0.3630
Resnet50	Si (52/126 congeladas)	Sí	100 y 200	20	0.9760	0.4059
Resnet50	Si (66/126 congeladas)	Sí	100 y 200	20	0.9659	0.3790
Resnet50	Si (38/126 congeladas)	Sí	100 y 200	20	0.9775	0.3850

**Tabla 4-4.** Tabla de resultados de CFEE 26 Emociones

La **Tabla 4-4** es la tabla de resultados de los experimentos realizados sobre las 26 clases de la base de datos Compound Facial Expressions of Emotion.

En este experimento nos hemos basado en los resultados de los anteriores antes de realizar las pruebas. En cuanto al modelo ResNet entrenado desde 0, hemos probado directamente con tamaños de Batch grandes. Como ocurría en los experimentos anteriores, los resultados con Batch = 64 y con Batch = 128 son similares, pero 128 con mayor *overfitting*.

Mejoramos los resultados con el modelo VGG16. En este hemos probado a utilizar y no utilizar Data Argumentation, obteniendo claramente un mayor *overfitting* cuando no la usábamos. La precisión de test se mantiene más o menos constante en todos los experimentos.

Los mejores resultados se siguen obteniendo con el modelo ResNet50. En este experimento el mejor resultado se obtuvo realizando un *Fine Tunning* hasta la capa 52 congelada. De esta manera se consigue obtener un 40,59 % de precisión. Es un muy buen resultado ya que si eligiésemos de una manera aleatoria la emoción, tendríamos una probabilidad de acierto del 3,85 % (1/26).

## 4.2 Resultados en detección de Parkinson

Los pacientes de Parkinson tienen una mayor dificultad que los pacientes sanos para controlar su musculatura. Cuando le dice que sonría a una persona con EP, es tan importante la imagen en la que se queda sonriendo como el proceso previo hasta que lo consigue. Es por esta razón por la que se utilizan las secuencias de imágenes mencionadas en el apartado 3.3.5 para el entrenamiento.

El primer experimento es en el cual generamos un modelo a partir de todas las capas menos la última del modelo que mejor resultado nos haya dado en la clasificación de CFEE, añadiéndole un clasificador de dos salidas que nos permitirá realizar la clasificación entre pacientes sanos y pacientes con EP. Se aplica la técnica de **Transfer Learning [3.2.1]**, aprovechando las características aprendidas en los experimentos de clasificación de emociones y aplicándolas para la tarea de clasificación entre pacientes sanos y pacientes con EP. Los resultados de clasificación son los que encontramos en la **Tabla 4-5**.

<b>Secuencia</b>	<b>Accuracy:</b>	<b>Sensitivity:</b>	<b>Specificity:</b>	<b>F1 Score:</b>	<b>EER:</b>
<b>SC-1</b>	$71.6 \pm 5.3$	$81.7 \pm 15.6$	$59.7 \pm 11.7$	$66.5 \pm 6.1$	$29.3 \pm 6.0$
<b>SC-2</b>	$71.4 \pm 7.1$	$86.1 \pm 9.8$	$53.5 \pm 10.2$	$65.1 \pm 8.7$	$31.7 \pm 9.0$
<b>SC-3</b>	$76.5 \pm 9.9$	$80.8 \pm 11.9$	$71.5 \pm 11.3$	$75.3 \pm 10.3$	$25.7 \pm 9.9$

**Tabla 4-5.** Tabla de resultados de 256 características.

Los mejores resultados se obtienen con la secuencia 3. Esta es la secuencia que incluye tanto el paso de neutral a sonriendo como de sonriendo a neutral. Al no actuar los mismos músculos en la contracción que en la relajación de la cara, el modelo aprenderá un mayor número de características con la SC-3 que con las otras dos, lo que le permitirá tener una mayor precisión a la hora de distinguir entre un paciente sano y un paciente con EP.

Para refinar un poco más los resultados, vamos a descongelar la penúltima capa. De esta manera tendremos congelados los pesos del modelo de clasificación CFEE hasta la capa de 2048 características, en vez de hasta la capa 256 como teníamos en el experimento anterior. Esto permitirá que el nuevo modelo aprenda un mayor número de características relacionadas con este nuevo dominio. Los resultados de este segundo experimento son los que podemos ver en la **Tabla 4-6**.

<b>Secuencia</b>	<b>Accuracy:</b>	<b>Sensitivity:</b>	<b>Specificity:</b>	<b>F1 Score:</b>	<b>EER:</b>
<b>SC-1</b>	$79.6 \pm 4.8$	$91.2 \pm 2.6$	$65.5 \pm 10.7$	$75.7 \pm 7.3$	$23.1 \pm 7.0$
<b>SC-2</b>	$76.3 \pm 6.1$	$88.0 \pm 6.3$	$62.2 \pm 12.8$	$71.9 \pm 9.0$	$21.2 \pm 3.6$
<b>SC-3</b>	$80.0 \pm 3.9$	$91.7 \pm 2.2$	$66.1 \pm 9.6$	$76.3 \pm 6.7$	$16.9 \pm 2.9$

**Tabla 4-6.** Tabla de resultados de 2048 características

Como era de esperar, los resultados mejoran. Esto se puede observar viendo como disminuye el EER (*Equal Error Rate*) en cualquiera de las secuencias. Las pruebas que mejores resultados dan siguen siendo las de la SC-3, por las mismas razones que en el experimento anterior. Los resultados que se obtienen son bastante buenos, consiguiendo una precisión del 80% en la clasificación entre pacientes sanos o enfermos.

## 5 Conclusiones y trabajo futuro

---

### 5.1 Conclusiones

Hoy en día las tecnologías relacionadas con el aprendizaje profundo son una realidad y han roto techos de cristal en áreas como la Visión por Computador. Son muchísimas las tareas para lo que es utilizado, entre las que se encuentran las relacionadas con el análisis facial. Dentro del análisis facial, muchos investigadores han trabajado con el dominio afectivo, es decir, con todo lo relacionado con el estudio de emociones. El estudio de los gestos faciales mediante redes neuronales entrenadas en dominios afectivos puede ayudar en el avance de ciertos campos, como por ejemplo en la medicina.

Un sector de la medicina en el que estas tecnologías pueden ayudar es en la detección de enfermedades neurodegenerativas. En este Trabajo de Fin de Grado nos hemos centrado, de entre todas las enfermedades neurodegenerativas, en el Parkinson. Realizando un estudio del estado del arte sobre el sector nos dimos cuenta que existen varias tecnologías computacionales que ayudan a una detección temprana de la Enfermedad de Parkinson, pero la mayoría de estas, están relacionadas con tecnologías del habla o escritura. Debido a esto pensamos en desarrollar un modelo de aprendizaje profundo que facilitase y ayudase en la tarea de detección de Parkinson a partir del estudio de los dominios afectivos en pacientes relacionados con la enfermedad.

Para la realización de este proyecto, los experimentos se basaron en dos fases. La primera fase consistía en realizar una clasificación de bases de datos de emociones a partir de redes neuronales pre-entrenadas y sin pre-entrenar. El objetivo de este experimento era buscar el modelo que mejor clasificase las emociones, de tal manera que aprendiese la mayor cantidad de características posibles del dominio afectivo, y se usase posteriormente para la detección de Parkinson. Con este experimento nos dimos cuenta que es mejor utilizar un modelo pre-entrenado que un modelo entrenado desde 0, ya que utiliza características aprendidas previamente para tu problema. Los modelos pre-entrenados que mejores resultados nos dieron fueron aquellos que habían sido entrenados en dominios parecidos a los del experimento actual. Esto es algo lógico ya que ambos modelos trabajan con características similares. También nos dimos cuenta que cuanto más controlado es el entorno (en cuanto a pose, iluminación...), mejores resultados se obtienen. Esto lo vimos al realizar experimentos con base de datos controladas y bases de datos *in to the wild*. Utilizando la técnica de Transfer Learning [3.2.1] cogimos el modelo que mejor resultado nos dio en la clasificación de emociones y le añadimos un clasificador de dos salidas. Con este nuevo modelo realizamos la segunda fase de experimentos, los relacionados con la detección de Parkinson. Estos experimentos se basaban en la clasificación de una base de datos con pacientes con EP. En estos experimentos conseguimos obtener un 80% de precisión a la hora de detectar si un paciente está sano o tiene Enfermedad de Parkinson, lo que puede ayudar, que nunca sustituir, a la detección médica de la Enfermedad de Parkinson. Gracias a este experimento podemos ver la importancia del Transfer Learning, ya que gracias

a las características aprendidas con el primer experimento podemos, sin más complicación que añadir un clasificador, realizar una tarea tan compleja como la detección de la Enfermedad de Parkinson a partir de ciertas imágenes. Esto nos da una idea de lo importante y buena que es esta técnica.

Finalmente quiero destacar la importancia que tiene que la tecnología y la medicina caminen de la mano. La tecnología computacional nunca va a sustituir a la medicina clásica, pero sí que la puede reforzar y ayudar a agilizar procesos. Por ejemplo, el modelo desarrollado en este Trabajo de Fin de Grado no va a aportar una verdad absoluta, ni va a ser capaz de sustituir a un neurólogo a la hora de detectar la Enfermedad de Parkinson, pero lo que sí que puede hacer es reforzar o refutar la decisión del médico. En caso de corroborarla el médico tendrá una mayor seguridad acerca de su decisión y en caso de refutarla, el médico podrá realizar una segunda consulta más en profundidad sobre el paciente. Por lo tanto la inclusión de este tipo de tecnologías en el área médica es algo muy positivo de cara al futuro.

## **5.2 Trabajo futuro**

Tras observar los resultados obtenidos en el trabajo, como trabajo futuro, nos gustaría mejorarlos en la medida de lo posible. Es cierto que un 80% es un muy buen resultado, pero si esta precisión en la clasificación se consiguiese aumentar a un 90 o 95%, el refuerzo a la medicina sería mucho mayor.

Para la mejora de estos porcentajes se requiere principalmente una mayor base de datos de pacientes con Enfermedad de Parkinson. El 80% ha sido alcanzado con una base de datos con 31 pacientes. Si esta base de datos se aumentase a miles de pacientes los resultados serían mucho mejores, ya que nuestros modelos serían capaces de aprender un mayor número de características.

Como trabajo futuro también se propone el desarrollo de una aplicación para móvil que sea capaz de realizar una detección temprana de Parkinson a partir de ciertos *selfies* y secuencias tomadas por el propio usuario de la aplicación.

## 6 Referencias

---

### 6.1 Referencias bibliográficas

- [1] R. Cacabelos, « Parkinson's disease: from pathogenesis to pharmaco genomics,» , International Journal of Molecular Sciences, 2017.
- [2] Matteo Bologna, Giovanni Fabbrini, Luca Marsili, Giovanni Defazio, Philip D Thompson, and Alfredo Berardelli, «Facial bradykinesia.,», Journal of Neurology, Neurosurgery and Psychiatry, 2013.
- [3] P. Ekman, «Strong evidence for universals in facial expressions: a reply to russell's mistaken critique.,», Psychological Bolletin, 1994.
- [4] . J. R. Orozco-Aroyave, J. C. Vásquez-Correa, J. F. Vargas-Bonilla, R. Arora, N. Dehak, P. S. Nidadavolu, H. Christensen, F. Rudzicz, M. Yancheva y H. Chinaei, «Neurospeech: An open-source software for parkinson's speech analysis,», Digital Signal Processing (Volume 77) , 2018.
- [5] C. G. Goetz, B. C. Tilley, . S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martínez Martin, C. Sampaio, W. Poewe, M. B. Stern y R. Dodel, «Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results.,» Movement Disorders (Volume 23), 2008.
- [6] H. Gassner, S. Steib, S. Klamroth, C. Federico , W. Adler, . B. Eskofier, K. Pfeifer, . J. Winkler y J. Klucken, « Perturbation Treadmill Training Improves Clinical Characteristics of Gait and Balance in Parkinson's Disease.,», Journal of Parkinson's Disease, 2019.
- [7] C. Rios-Urrego, J. C. V'asquez-Correa, J. F. Vargas-Bonilla, E. Nöth, F. Lopera y J. R. Orozco-Aroyave, « Analysis and evaluation of handwriting in patients with parkinson's disease using kinematic, geometrical, and non-linear features,», Computer Methods and Programs in Biomedicine, (Volume 173) 2019.
- [8] S. Spasojevic, T. V. Ilic, I. Stojkovic, V. Potkonjak, A. Rodic y J. Santos-Victor, « Quantitative assessment of the arm/hand movements in parkinson's disease using a wireless armband device.,» , Frontiers in Neurology, 2017.
- [9] A. Bandini, S. Orlandi, H. J. Escalante, F. Giovannelli, M. Cincotta, C. A. Reyes-Garcia, P. Vanni, G. Zaccara y C. Manfredi, «Analysis of facial expressions in parkinson's disease through video-based automatic methods,», Journal of neurosciences methods, 2017.
- [10] S. Argaud, M. Vérin, P. Sauleau y D. Grandjean, «Facial emotion recognition in parkinson's disease: A review and new hypotheses,», Movement disorders: official journal of the Movement Disorde Soociety, 2018.
- [11] O. Celiktutan, E. Skordos y H. Gunes, « Multimodal human-human-robot interactions (mhhri) dataset for studying personality and engagement,», IEEE Transactions on Affective Computing (Volume 10), 2017.
- [12] K. Li, W. Tao y L. Liu, « Online semantic object segmentation for vision robot collected video.,», IEEE Access (Volume 7), 2019.
- [13] A. Parnandi y R. Gutierrez-Osuna, «Visual biofeedback and game adaptation in relaxation skill transfer,», ,», IEEE Transactions on Affective Computing (Volume 10), 2017.



- [14] W. Peng, I. Gonzalez, G. Patsis, D. Jiang, H. Sahli, E. Kerckhofs y M. Vandekerckhove, « Objectifying facial expressivity assessment of parkinson's patients: preliminary study.», Computational and mathematical methods in medicine, 2014.
- [15] J. Hamm , C. G. Kohler, R. C. Gur y R. Verma, «Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders», Journal of Neuroscience Methods, 2011.
- [16] A. Pampouchidou, P. Simos, K. Marias, F. Meriaudeau, F. Yang, M. Pediaditis y M. Tsiknakis, « Automatic assessment of depression based on visual cues: A systematic review», IEEE Transactions on Affective Computing (Volume 10), 2017.
- [17] L. Gómez-Gómez, A. Morales y J. Orozco-Arroyave, «Affective domain adaptation with transfer learning in CNN for detection of Parkinson's patients.», Presented at 2019 Prognostics and System Health Management Conference (PHM 2019), 2020.
- [18] W. Zhao, R. Chellapa, P. Phillips y A. Rosenfeld, «Face recognition: a literature survey», ACM Computing Surveys, (Volume 35), 2003.
- [19] A. Krizhevsky, I. Sutskever y E. Hinton, «Imagenet classification with deep convolutional neural networks», Advances in neural information processing systems (Volume 25), 2012.
- [20] K. He, X. Zhang, S. Ren y J. Sun, «Deep residual learning for image recognition», Conference on Computer Vision and Pattern Recognition, 2016.
- [21] R. Girshick, J. Donahue, T. Darrell y J. Malik, «Rich feature hierarchies for accurate object detection and semantic segmentation», Conference on Computer Vision and Pattern Recognition, 2014.
- [22] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu y A. Berg, «SSD: Single shot multibox detector», European Conference on Computer Vision, 2016.
- [23] R. Ranjan, S. Sankaranarayanan, C. Castillo y R. Chellappa, «An all-in-one convolutional neural network for face analysis», International Conference on Automatic Face and Gesture Recognition, 2017.
- [24] R. Ranjan, V. Patel y R. Chellappa, «A deep pyramid deformable part model for face detection», IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2015
- [25] R. Ranjan, V. Patel y R. Chellappa, «“Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition», IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.
- [26] A. Kumar, R. Ranjan, V. Patel y R. Chellappa, «Face alignment by local deep descriptor regression», IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016.
- [27] D. Yi, Z. Lei, S. Liao y S. Z. Li, «Learning face representation from scratch», 2014.
- [28] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller y E. Brossard, «The megaface benchmark: 1 million faces for recognition at scale», Conference: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [29] A. Nech y I. Kemelmacher-Shlizerman, «Level playing field for million scale face recognition», 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [30] Y. Guo, L. Zhang, Y. Hu, X. He y J. Gao, «Ms-celeb-1m: A data set and benchmark for large-scale face recognition», The 14th European Conference on Computer

Vision, 2016.

- [31] O. M. Parkhi, A. Vedaldi y A. Zisserman, «Deep face recognition», British Machine Vision Conference 2015, 2015.
- [32] M. Pantic y L. J. Rothkrantz, «Expert system for automatic analysis of facial expressions», 2000.
- [33] Z. Li, J. -i. Imai y M. Kaneko, «Facial-component-based bag of words and phog descriptor for facial expression recognition», Conference: Systems, Man and Cybernetics, 2009.
- [34] E. Friesenand y P. Ekman, «Facial action coding system: a technique for the measurement of facial movement», 1978.
- [35] P. Ekman y W. V. Friesen, «Constants across cultures in the face and emotion», Journal of personality and social psychology, 1971.
- [36] M. Soleymani, S. Asghari-Esfeden, Y. Fu y M. Pantic, «Analysis of eeg signals and facial expressions for continuous emotion detection», IEEE Transactions on Affective Computing (Volume 7), 2016.
- [37] S. Du, Y. Tao y A. M. Martinez, «Compound facial expressions of emotion», Proceedings of the National Academy of Sciences (Volume 111), 2014.
- [38] M. A. Nicolaou, H. Gunes y M. Pantic, «Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space», IEEE Transactions on Affective Computing (Volume 2), 2011.
- [39] K. Schindler, L. Van Gool y B. de Gelder, «Recognizing emotions expressed by body pose: A biologically inspired neural model», Neural Networks (Volume 21), 2008.
- [40] W. Mou, O. Celiktutan y H. Gunes, «Group-level arousal and valence recognition in static images: Face, body and context», IEEE International Conference on Automatic Face and Gesture Recognition, 2015.
- [41] E. Dorsey, R. Constantinescu, J. P. Thompson, K. M. Biglan, R. G. Holloway, K. Kiebertz, F. J. Marshall, B. M. Ravina, G. Schifitto, A. Siderowf y C. M. Tanner, «Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030», Neurology (Volume 69) 2007.
- [42] C. M. Tanner, M. Brandabur y E. R. Dorsey, « Parkinson Disease: A Global View», Journal of Parkinson Disease, 2008.
- [43] T. Pitcairn , S. Clemie , J. Gray y B. Pentland , «Non-verbal cues in the self-presentation of Parkinsonian patients», The British journal of clinical psychology, 1990.
- [44] M. Smith , M. Smith y H. Ellgring , « Spontaneous and posed facial expression in Parkinson's Disease», Cambridge University Press, 1996.
- [45] M. Frank , P. Ekman y W. Friesen , « Behavioral markers and recognizability of the smile of enjoyment.», Journal of Personality and Social Psychology, 1993.
- [46] S. Gunnery y M. Ruben , «Perceptions of Duchenne and non-Duchenne smiles: A meta-analysis.», Frontiers in Psychology, 2016.
- [47] M. Bologna , G. Fabbrini , L. Marsili , G. Defazio , P. Thompson y A. Berardelli , «Facial bradykinesia», Journal of Neurology, Neurosurgery and Psychiatry, 2013.
- [48] D. Bowers , K. Miller , W. Bosch , D. Gokcay , O. Pedraza , U. Springer y M. Okun , « Faces of emotion in Parkinsons disease: Micro-expressivity and bradykinesia during voluntary facial expressions», Journal of the International Neuropsychological Society, 2006.

- [49] K. Dujardin , « Deficits in decoding emotional facial expressions in Parkinson's disease,», *Neuropsychologia* (Volume 42), 2004.
- [50] N. Vinokurov , «Quantifying hypomimia in Parkinson patients using a depth camera,», from book *Pervasive Computing Paradigms for Mental Health*, 2015.
- [51] W. Rinn , « The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions,», *Psychological Bulletin* (Volume 95), 1984.
- [52] P. Wu , I. Gonzalez , G. Patsis , D. Jiang , H. Sahli , E. Kerckhofs y M. Vandekerckhove , « Objectifying facial expressivity assessment of Parkinson's patients: Preliminary study.,», *Computational and Mathematical Methods in Medicine*, 2014.
- [53] C. Goetz , B. Tilley , S. Shaftman , G. Stebbins , S. Fahn , P. Martinez-Martin y W. Poewe , «Movement disorder society-sponsored revision of the unified Parkinson's Disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results.,» *Movement Disorders* (Volume 23), 2008.
- [54] A. Calder , A. Young , J. Keane y M. Dean , «Configural information in facial expression perception.,», *Journal of Experimental Psychology Human Perception & Performance* (Volume 26), 2000.
- [55] M. Knapp, J. Hall y T. Horgan, « Nonverbal communication in human interaction,», Wadsworth Publishing, 2013.
- [56] G. Simons, M. Pasqualini, V. Reddy y J. Wood, « Emotional and nonemotional facial expressions in people with Parkinson's disease,», *Journal of the International Neuropsychological Society: JINS* (Volume 10), 2004.
- [57] N. Vinokurov, D. Arkadir, E. Linetsky, H. Bergman y D. Weinshall, « Quantifying Hypomimia in Parkinson Patients Using a Depth Camera,», from book *Pervasive Computing Paradigms for Mental Health*, 2015.
- [58] L. Ricciardi , P. Baggio , D. Ricciardi , B. Morabito , M. Pomponi , A. Bentivoglio y D. Volpe , « Rehabilitation of hypomimia in Parkinson's disease: a feasibility study of two different approaches,», *Neurological Sciences: official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 2015.
- [59] A. Grammatikopoulou, N. Grammalidis, S. Bostantjopoulou y Z. Katsarou , «Detecting Hypomimia Symptoms By Selfie Photo Analysis,», *International Conference on Pervasive Technologies Related to Assistive Environments*, 2019.
- [60] S. D. Gunnery, E. N. Naumova, M. Saint-Hilaire y L. Tickle-Degnen, «Mapping spontaneous facial expression in people with Parkinson's disease: A multiple case study design,», *Congent Psychology* (Volume 4), 2017.
- [61] R. Kosti, J. M. Alvarez, A. Recasens y A. Lapedriza, «Context Based Emotion Recognition using EMOTIC Dataset,», *IEEE transactions on pattern analysis and machine intelligence*, 2019.
- [62] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J.-C. Chen, V. M. Patel, C. D. Castillo y R. Chellappa, «Deep Learning for Understanding Faces,», *IEEE Signal Processing Magazine* (Volume 35), 2018.
- [63] A. Mollahosseini, B. Hasani y M. H. Mahoor, «AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild,», *IEEE Transactions on Affective Computing*, 2017.
- [64] F. Chollet, *Deep Learning with Python*, editorial Manning, 2018.
- [65] M. u. Hassan, «<https://neurohive.io/en/popular-networks/vgg16/>,» 2018. [En línea].

- [66] P. Jay, «<https://medium.com/@14prakash/understanding-and-implementing-architectures-of-resnet-and-resnext-for-state-of-the-art-image-cf51669e1624>,» 2018. [En línea].
- [67] K. He, X. Zhang, S. Ren y S. Jian , «<https://www.kaggle.com/keras/resnet50>,» 2017. [En línea].
- [68] Q. Cao, L. Shen, W. Xie, O. M. Parkhi y A. Zisserman, «VGGFace2: A dataset for recognising face across pose and age,», 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 2018.
- [69] D. Lazar, «<https://towardsdatascience.com/building-a-resnet-in-keras-e8f1322a49ba>,» 2020. [En línea].





